

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 13:37:44 : Search time 15.4854 Seconds
(without alignments)
1829.738 Million cell updates/sec

Title: US-10-073-333a-2
Perfect score: 963
Sequence: 1 MLFRARGPVRCRWGPAEA.....TCYFWKKKQKKKTLNLFN 963

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0
Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	0.9	1487	2	US-08-760-489-2
2	9	0.9	1487	2	US-08-760-489-4
3	9	0.9	1487	4	US-09-185-373-2
4	9	0.9	1487	4	US-09-185-373-4
5	8	0.8	100	1	US-08-158-189-7
6	8	0.8	133	4	US-08-468-560C-8
7	8	0.8	139	2	US-08-219-237B-8
8	8	0.8	140	4	US-08-477-347-17
9	8	0.8	140	4	US-08-476-862-8
10	8	0.8	153	4	US-09-286-529-2
11	8	0.8	205	3	US-08-974-022-51
12	8	0.8	205	4	US-08-795-445A-51
13	8	0.8	205	4	US-08-795-447A-51
14	8	0.8	205	4	US-08-974-186-51
15	8	0.8	205	4	US-08-795-446B-51
16	8	0.8	205	4	US-08-706-945D-138
17	8	0.8	206	1	US-08-097-827-7
18	8	0.8	206	1	US-08-494-574-7
19	8	0.8	211	4	US-09-286-529-20
20	8	0.8	299	4	US-03-286-529-17
21	8	0.8	300	2	US-08-794-796-2
22	8	0.8	438	1	US-08-097-827-11
23	8	0.8	438	1	US-08-494-574-11
24	7	0.7	15	4	US-08-602-999A-363
25	7	0.7	15	4	US-03-500-124-363
26	7	0.7	77	4	US-03-146-950-25
27	7	0.7	155	4	US-09-146-950-4

28	7	0.7	159	4	US-09-146-950-20	Sequence 20, Appl
29	7	0.7	193	4	US-09-146-950-2	Sequence 2, Appli
30	7	0.7	197	4	US-09-146-950-18	Sequence 18, Appl
31	7	0.7	278	2	US-08-569-168-2	Sequence 2, Appli
32	7	0.7	283	4	US-08-509-024-2	Sequence 2, Appli
33	7	0.7	283	4	US-09-333-279-2	Sequence 2, Appli
34	7	0.7	283	4	US-09-072-993C-2	Sequence 2, Appli
35	7	0.7	283	5	PCT-US96-12374-2	Sequence 2, Appli
36	7	0.7	320	2	US-08-933-750C-12	Sequence 12, Appl
37	7	0.7	320	4	US-09-234-613-12	Sequence 12, Appl
38	7	0.7	331	4	US-09-086-483A-3	Sequence 3, Appli
39	7	0.7	383	2	US-08-569-168-7	Sequence 7, Appli
40	7	0.7	419	4	US-08-509-024-7	Sequence 7, Appli
41	7	0.7	419	4	US-09-333-279-7	Sequence 7, Appli
42	7	0.7	479	4	US-09-177-349-3	Sequence 3, Appli
43	7	0.7	588	5	PCT-US95-13749-4	Sequence 4, Appli
44	7	0.7	761	1	US-07-906-395-2	Sequence 2, Appli
45	7	0.7	761	1	US-08-192-632-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-760-489-2
; Sequence 2, Application US/08760489
; Patent No. 5830696
; GENERAL INFORMATION:
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: DIRECTED EVOLUTION OF THERMOPHILIC
; TITLE OF INVENTION: ENZYMES
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: PastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/760,489
FILING DATE: 05-DEC-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/008,311
FILING DATE: 07-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Haile, Ph.D., Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/008001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1487 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: Internal
US-08-760-489-2

Query Match 0.9%; Score 9; DB 2; Length 1487;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 EAPRRGRSP 27
|||||||

```

DB 171 EAPRRGRSP 179
;
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/185.373
; FILING DATE: 03-NO. 6335179-1998
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/760.489
; FILING DATE: 05-DEC-1996
; APPLICATION NUMBER: 60/008.311
; FILING DATE: 07-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/008001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1487 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
;
; US-09-185-373-2
;
; Query Match 0.9%; Score 9; DB 4; Length 1487;
; Best Local Similarity 100.0%; Pred. No. 3.6;
; Matches 9; Conservative 0; Mismatches 0; Indels
;
Qy 19 EAPRRGRSP 27
| | | | | | | |
Db 171 EAPRRGRSP 179
;
; RESULT 4
; US-09-185-373-4
; Sequence 4, Application US/09185373
; Patent No. 6335179
; GENERAL INFORMATION:
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: DIRECTED EVOLUTION OF THERMOPHILIC
; ENZYMES
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/185.373
; FILING DATE: 03-NO. 6335179-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/760.489
; FILING DATE: 05-DEC-1996
; APPLICATION NUMBER: 60/008.311
; FILING DATE: 07-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/008001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1487 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
;
; US-08-760-489-4
;
; Query Match 0.9%; Score 9; DB 2; Length 1487;
; Best Local Similarity 100.0%; Pred. No. 3.6;
; Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
Qy 19 EAPRRGRSP 27
| | | | | | | |
Db 171 EAPRRGRSP 179
;
; RESULT 3
; US-09-185-373-2
; Sequence 2, Application US/09185373
; Patent No. 6335179
; GENERAL INFORMATION:
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: DIRECTED EVOLUTION OF THERMOPHILIC
; ENZYMES
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:

```

```

DB 171 EAPRRGRSP 179
;
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/185.373
; FILING DATE: 03-NO. 6335179-1998
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/760.489
; FILING DATE: 05-DEC-1996
; APPLICATION NUMBER: 60/008.311
; FILING DATE: 07-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/008001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1487 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
;
; US-09-185-373-2
;
; Query Match 0.9% Score 9; DB 4; Length 1487;
; Best Local Similarity 100.0%; Pred. No. 3.6;
; Matches 9; Conservative 0; Mismatches 0; Indels
;
QY 19 EAPRRGRSP 27
| | | | | | | |
DB 171 EAPRRGRSP 179
;
; RESULT 4
; US-09-185-373-4
; Sequence 4, Application US/09185373
; Patent No. 6335179
; GENERAL INFORMATION:
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: DIRECTED EVOLUTION OF THERMOPHILIC
; ENZYMES
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/185.373
; FILING DATE: 03-NO. 6335179-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/760.489
; FILING DATE: 05-DEC-1996
; APPLICATION NUMBER: 60/008.311
; FILING DATE: 07-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/008001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1487 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
;
; US-08-760-489-4
;
; Query Match 0.9% Score 9; DB 2; Length 1487;
; Best Local Similarity 100.0%; Pred. No. 3.6;
; Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 19 EAPRRGRSP 27
| | | | | | | |
DB 171 EAPRRGRSP 179
;
; RESULT 3
; US-09-185-373-2
; Sequence 2, Application US/09185373
; Patent No. 6335179
; GENERAL INFORMATION:
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: DIRECTED EVOLUTION OF THERMOPHILIC
; ENZYMES
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1487 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Query Match 0.9%; Score 9; DB 4; Length 1487;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 EAPRRGRSP 27
| | | | | | | | | |
Db 171 EAPRRGRSP 179

RESULT 5
US-08-158-189-7
; Sequence 7, Application US/08158189
; Patent No. 5641497
; GENERAL INFORMATION:
; APPLICANT: Bevins, Charles L.
; APPLICANT: Jones, Douglas E.
; TITLE OF INVENTION: Gastrointestinal Defensin Peptides,
; TITLE OF INVENTION: cDNA Sequences, Methods for Production and Use Thereof
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz and No. 5641497ris
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/158.189
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/888,232
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Johnson, Philip S.
; REGISTRATION NUMBER: 27,200
; REFERENCE/DOCKET NUMBER: CH-0219
; TELEPHONE: 215-568-3100
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-158-189-7

Query Match 0.8%; Score 8; DB 1; Length 100;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 936 TAVLLVAL 943
| | | | | | | | | |
Db 8 TAVLLVAL 15

RESULT 6
US-08-468-560C-8
; Sequence 8, Application US/08468560C
; Patent No. 6270998
; GENERAL INFORMATION:
; APPLICANT: NAGATA, Shigekazu
; APPLICANT: ITOH, Naoto
; APPLICANT: YONEHARA, Shin
; TITLE OF INVENTION: DNA CODING FOR HUMAN CELL SURFACE
; TITLE OF INVENTION: ANTIGEN
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP.
; STREET: P.O. BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468.560C
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY JR., GERLAD M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 20-4393P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 133 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-468-560C-8

Query Match 0.8%; Score 8; DB 4; Length 133;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 616 CVPCPGGH 623
| | | | | | | | | |
Db 91 CVPCPGGH 98

RESULT 7
US-08-219-237B-8
; Sequence 8, Application US/08219237B
; Patent No. 5874546
; GENERAL INFORMATION:
; APPLICANT: NAGATA, Shigekazu
; APPLICANT: ITOH, Naoto
; APPLICANT: YONEHARA, Shin
; TITLE OF INVENTION: DNA Coding for Human Cell Surface Antigen
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: James W. Hellwege
; STREET: P.O. Box 2266 Eads Station
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/219,237B
FILING DATE: 28-MAR-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/872,129
FILING DATE: 22-APR-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: James W. Hellwege
REGISTRATION NUMBER: 28,808
REFERENCE/DOCKET NUMBER: 516762
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 139 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-219-237B-8

Query Match 0.8%; Score 8; DB 2; Length 139;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 616 CVCPPPGH 623
Db 97 CVCPPPGH 104

RESULT 8
US-08-477-347-17
Sequence 17, Application US/08477347
Patent No. 6232446
GENERAL INFORMATION:
APPLICANT: WALLACH, David
APPLICANT: BIGDA, Jacek
APPLICANT: BELETSKY, Igor
APPLICANT: METT, Igor
TITLE OF INVENTION: TNF LIGANDS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,347
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/115,685
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 106271
FILING DATE: 08-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Townsend, G. Kevin
REGISTRATION NUMBER: 34,033
REFERENCE/DOCKET NUMBER: WALLACH=10
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633

INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 140 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-477-347-17

Query Match 0.8%; Score 8; DB 4; Length 140;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 616 CVCPPPGH 623
Db 98 CVCPPPGH 105

RESULT 9
US-08-476-862-8
Sequence 8, Application US/08476862
Patent No. 6262239
GENERAL INFORMATION:
APPLICANT: WALLACH, David
APPLICANT: BIGDA, Jacek
APPLICANT: BELETSKY, Igor
APPLICANT: METT, Igor
APPLICANT: ENGELMANN, Hartmut
TITLE OF INVENTION: TNF INHIBITORS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/476,862
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 107267
FILING DATE: 12-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 94039
FILING DATE: 06-APR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 91229
FILING DATE: 06-AUG-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 90339
FILING DATE: 18-MAY-1989
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: WALLACH=12A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 140 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-476-862-8

Query Match 0.8%; Score 8; DB 4; Length 140;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 616 CVPCPGGH 623
Db 98 CVPCPGGH 105
|||||

RESULT 10
US-09-286-529-2
; Sequence 2, Application US/09286529
; Patent No. 6257367
; GENERAL INFORMATION:
; APPLICANT: Catherine Tribouley
; TITLE OF INVENTION: NEW MEMBERS OF TNF AND TNFR FAMILIES
; FILE REFERENCE: 1408.003/200130.439C1
; CURRENT APPLICATION NUMBER: US/09/286,529
; CURRENT FILING DATE: 1999-04-05
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 153
; TYPE: PRT
; ORGANISM: human
US-09-286-529-2

Query Match 0.8%; Score 8; DB 4; Length 153;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 410 PCPPGTFS 417
Db 68 PCPPGTFS 75
|||||

RESULT 11
US-08-974-022-51
; Sequence 51, Application US/08974022
; Patent No. 6015938
; GENERAL INFORMATION:
; APPLICANT: Boyle, William J.
; APPLICANT: Lacey, David L.
; APPLICANT: Calzone, Frank J.
; APPLICANT: Chang, Ming-Shi
; TITLE OF INVENTION: OSTEOPROTEGERIN
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,022
; FILING DATE: 12-DEC-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/577,788
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-378
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 205 amino acids
; TYPE: amino acids
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-974-022-51

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-974-022-51

Query Match 0.8%; Score 8; DB 3; Length 205;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 616 CVPCPGGH 623
Db 122 CVPCPGGH 129
|||||

RESULT 12
US-08-795-445A-51
; Sequence 51, Application US/08795445A
; Patent No. 6284485
; GENERAL INFORMATION:
; APPLICANT: Boyle, William J.
; APPLICANT: Lacey, David L.
; APPLICANT: Calzone, Frank J.
; APPLICANT: Chang, Ming-Shi
; TITLE OF INVENTION: OSTEOPROTEGERIN
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,445A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/577,788
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-378
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 205 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-795-445A-51

Query Match 0.8%; Score 8; DB 4; Length 205;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 616 CVPCPGGH 623
Db 122 CVPCPGGH 129
|||||

RESULT 13
US-08-795-447A-51
; Sequence 51, Application US/08795447A
; Patent No. 6284728
; GENERAL INFORMATION:
; APPLICANT: Boyle, William J.
; APPLICANT: Lacey, David L.

APPLICANT: Calzone, Frank J.
APPLICANT: Chang, Ming-Shi
TITLE OF INVENTION: Osteoprotegerin
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen Inc.
STREET: One Amgen Center Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91362-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/795,447A
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Winter, Robert B.
REFERENCE/DOCKET NUMBER: A-378D2
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 205 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-795-447A-51

Query Match 0.8%; Score 8; DB 4; Length 205;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 616 CVPCCPGH 623
Db 122 CVPCCPGH 129

RESULT 14
US-08-974-186-51
Sequence 51, Application US/08974186
Patent No. 6284740
GENERAL INFORMATION:
APPLICANT: Boyle, William J.
APPLICANT: Lacey, David L.
APPLICANT: Calzone, Frank J.
APPLICANT: Chang, Ming-Shi
TITLE OF INVENTION: OSTEOPROTEGERIN
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen Inc.
STREET: 1840 Dehavilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974.186
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/577,788
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Winter, Robert B.

REFERENCE/DOCKET NUMBER: A-378
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 205 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-974-186-51

Query Match 0.8%; Score 8; DB 4; Length 205;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 616 CVPCCPGH 623
Db 122 CVPCCPGH 129

RESULT 15
US-08-795-446B-51
Sequence 51, Application US/08795446B
Patent No. 6288032
GENERAL INFORMATION:
APPLICANT: Boyle, William J.
APPLICANT: Lacey, David L.
APPLICANT: Calzone, Frank J.
APPLICANT: Chang, Ming-Shi
TITLE OF INVENTION: OSTEOPROTEGERIN
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen Inc.
STREET: 1840 Dehavilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/795.446B
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/577,788
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Winter, Robert B.
REFERENCE/DOCKET NUMBER: A-378
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 205 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-795-446B-51

Query Match 0.8%; Score 8; DB 4; Length 205;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 616 CVPCCPGH 623
Db 122 CVPCCPGH 129

Search completed: May 12, 2003, 13:42:22
Job time : 17.4854 secs

GenCore version 5.1.5
 Copyright (c) 1993 - 2003 Compugen Ltd.
 OM protein - protein search, using sw model
 Run on: May 12, 2003, 13:30:09 ; Search time 31.4548 Seconds
 (without alignments)
 6308.212 Million cell updates/sec
 Title: US-10-073-333A-2
 Perfect score: 963
 Sequence: 1 MLFRARGPVRGNGRPAEA.....TCYFWKKNQKKKTLTLNLFN 963

Scoring table: OLIGO
 Gapop 60.0 , Gapext 60.0
 Searched: 671580 seqs, 206047115 residues
 Word size : 0
 Total number of hits satisfying chosen parameters: 671580
 Minimum DB seq length: 0
 Maximum DB seq length: 2000000000
 Post-processing: Listing first 45 summaries

Database : SPTREMBL_21.*
 1: sp_archaea.*
 2: sp_bacteria.*
 3: sp_fungi.*
 4: sp_human.*
 5: sp_invertebrate.*
 6: sp_mammal.*
 7: sp_mhc.*
 8: sp_organelle.*
 9: sp_phage.*
 10: sp_plant.*
 11: sp_rodent.*
 12: sp_virus.*
 13: sp_vertebrate.*
 14: sp_unclassified.*
 15: sp_ivirus.*
 16: sp_bacteriaph.*
 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	342	35.5	493	Q96DP2	Q96dp2 homo sapien
2	11	1.1	300	11 Q8R215	Q8r215 mus musculu
3	11	1.1	580	4 Q9P2W2	Q9p2w2 homo sapien
4	9	0.9	552	10 Q82444	Q82444 nicotiana t
5	8	0.8	55	12 Q67608	Q67608 tomato gold
6	8	0.8	148	17 Q9HR77	Q9hr77 halobacteri
7	8	0.8	154	12 Q9DGZ5	Q9dgz5 hepatitis b
8	8	0.8	154	12 Q9E6T5	Q9e6t5 hepatitis b
9	8	0.8	239	16 Q9ABT4	Q9abt4 caulobacter
10	8	0.8	268	16 Q92SW3	Q92sw3 listeria mo
11	8	0.8	281	5 Q9TZN3	Q9tzn3 caenorhabdi
12	8	0.8	286	16 Q9PR14	Q9pr14 ureaplasma
13	8	0.8	293	12 Q8V911	Q8v911 squash yell
14	8	0.8	379	2 Q50198	Q50198 mycobacteri
15	8	0.8	380	8 Q9TFK5	Q9tfr5 etheostoma
16					

17	8	0.8	380	8	Q958X6	Q958x6 etheostoma
18	8	0.8	428	16	Q8ZF14	Q8zf14 yersinia pe
19	8	0.8	444	10	Q23079	Q23079 arabidopsis
20	8	0.8	447	5	O44593	O44593 caenorhabdi
21	8	0.8	498	4	O43267	O43267 homo sapien
22	8	0.8	550	4	Q9P2Q1	Q9p2q1 homo sapien
23	8	0.8	593	10	Q8RUU6	Q8ruu6 zea mays (m
24	8	0.8	597	10	Q9LGL5	Q9lgl5 oryza sativ
25	8	0.8	605	10	Q8S4W8	Q8s4w8 zea mays (m
26	8	0.8	606	10	Q8S4W9	Q8s4w9 zea mays (m
27	8	0.8	607	10	Q9FFT4	Q9fft4 arabidopsis
28	8	0.8	607	10	Q96536	Q96536 arabidopsis
29	8	0.8	622	2	Q45155	Q45155 bacteroides
30	8	0.8	704	10	Q04434	Q04434 pisum sativ
31	8	0.8	721	11	Q922S8	Q922s8 mus musculu
32	8	0.8	742	16	Q8YNZ6	Q8ynz6 anabaena sp
33	8	0.8	881	10	Q65924	Q65924 arabidopsis
34	8	0.8	1051	12	Q91EU4	Q91eu4 cydia pomon
35	8	0.8	1059	10	Q9LH84	Q9lhs4 arabidopsis
36	8	0.8	1206	16	Q9CCX9	Q9ccx9 mycobacteri
37	8	0.8	1232	10	Q9LJQ1	Q9ljq1 arabidopsis
38	8	0.8	1578	16	Q92E25	Q92e25 listeria in
39	8	0.8	1582	16	Q8Y9A5	Q8y9a5 listeria mo
40	8	0.8	2295	5	Q9TY98	Q9ty98 plasmodium
41	8	0.8	2712	16	Q9F3X5	Q9f3x5 pasteurella
42	8	0.8	3102	5	Q9TZR4	Q9tzt4 caenorhabdi
43	8	0.8	3503	5	Q24292	Q24292 drosophila
44	7	0.7	65	6	Q95J37	Q95j37 sus scrofa
45	7	0.7	79	2	Q9AFQ6	Q9afq6 shigella fl

ALIGNMENTS

RESULT 1

Q96DP2: PRELIMINARY; PRT; 493 AA.
 AC Q96DP2;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE CDNA FLJ31340 fis, clone MESAN100035, weakly similar to major surface-labeled trophozoite antigen precursor.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
 RT "NEO human cDNA sequencing project."
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK055902; BAB71041.1;
 DR InterPro; IPR001881; EGF_Ca.
 DR SMART; SM00179; EGF_CA; 1.
 SQ SEQUENCE 493 AA; 54585 MW; 0AEBB2C19709B211 CRC64;

Query Match 35.5%; Score 342; DB 4; Length 493;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 342; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 334 GSSECTERPPCTTKDYFQIHPCDEEGTKQIMYKWIPEKICREDLTDAIRLPPSGEKKDC 393
 Db 96 GSSECTERPPCTTKDYFQIHPCDEEGTKQIMYKWIPEKICREDLTDAIRLPPSGEKKDC 155
 Qy 394 PPCNPGFYNNGSSSCHPCPPGTFSDGTKECRPCPAGTEPALGFEYKWNVLPGNMKTSCF 453

Db 156 PPCNPGYNNSSCHPCPPCTGDKECRCPCAGTEPALGFEYKWNVLPGNMKTSF 215
|||||
Qy 454 NVGNSKCDGMNGWVAGDHTQSGAGGSDNDYLLNLHLIPGFKPPTSMGTGATGSELGRITF 513
|||||
Db 216 NVGNSKCDGMNGWVAGDHTQSGAGGSDNDYLLNLHLIPGFKPPTSMGTGATGSELGRITF 275
|||||
Qy 514 VFETLCSADCVLYFMVDINRSTNVVSWGGTKEKQAYTHIIFKNATFTFTWAFQRTNOG 573
|||||
Db 276 VFETLCSADCVLYFMVDINRSTNVVSWGGTKEKQAYTHIIFKNATFTFTWAFQRTNOG 335
|||||
Qy 574 QDNRRFINDMKYIYSITATNAVGVASSCRACALGSESGSCVPCPPGHVIERETNOCK 633
|||||
Db 336 QDNRRFINDMKYIYSITATNAVGVASSCRACALGSESGSCVPCPPGHVIERETNOCK 395
|||||
Qy 634 ECPDPTYLISHQVYGEACICPGGSKNNQDHSVCYSDCFY 675
|||||
Db 396 ECPDPTYLISHQVYGEACICPGGSKNNQDHSVCYSDCFY 437
|||||
RESULT 2
Q8R215 PRELIMINARY; PRT; 300 AA.
AC Q8R215;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 32.6 kDa protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Strausberg R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC022655; AAB22655.1; -
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 300 AA; 32642 MW; 6455109C054B6CB4 CRC64;
Query Match 1.1%; Score 11; DB 11; Length 300;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 943 LTCYFWKKKQK 953
|||||
Db 213 LTCYFWKKKQK 223
|||||
RESULT 3
Q9P2M2 PRELIMINARY; PRT; 580 AA.
AC Q9P2M2;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE KIAA1324 protein (Fragment).
GN KIAA1324.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=20181126; PubMed=10718198;
RA Nagase T., Kikuno R., Ishikawa K., Hirosewa M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.XVI.
RT The complete sequences of 150 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:65-73(2000).

DR EMBL; AB037745; BAA92562.1; -
FT NON_TER 1
SQ SEQUENCE 580 AA; 63365 MW; 5E63C19265EC8E5C CRC64;
Query Match 1.1%; Score 11; DB 4; Length 580;
Best Local Similarity 100.0%; Pred. No. 0.027;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 510 RITEVFETLCS 520
|||||
Db 72 RITEVFETLCS 82
|||||
RESULT 4
O82444 PRELIMINARY; PRT; 552 AA.
ID O82444;
AC O82444;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Peroxisomal targeting sequence 1 receptor (Fragment).
GN PEX5.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95007315; PubMed=9789089;
RA Kragler F., Lametschwandner G., Christmann J., Hartig A.,
RA Harada J.;
RT "Identification and analysis of the plant peroxisomal targeting signal
RL 1 receptor NPEX5.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:13336-13341(1998).
DR EMBL; AF056282; AAC69180.1; -
DR InterPro; IPR001440; TPR.
DR Pfam; PF00513; TPR; 4.
DR SMART; SM00028; TPR; 4.
KW Receptor.
FT NON_TER 1
SQ SEQUENCE 552 AA; 62293 MW; 7045FA177B0F51C6 CRC64;
Query Match 0.9%; Score 9; DB 10; Length 552;
Best Local Similarity 100.0%; Pred. No. 3.4; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 4 RARGPVRGR 12
|||||
Db 22 RARGPVRGR 30
|||||
RESULT 5
Q67608 PRELIMINARY; PRT; 55 AA.
ID Q67608;
AC Q67608;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Tomato golden mosaic virus subgenomic DNA derived from DNA B cccds -
DE covalently closed circular double-stranded molecule.
OS Tomato golden mosaic virus (TMGV).
OC Viruses; ssDNA viruses; Geminiviridae; Begomovirus.
OX NCBI_TaxID=10831;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87040767; PubMed=3022243;
RA MacDowell S.W., Coutts R.H.A., Buck K.W.;
RT "Molecular characterization of subgenomic single-stranded and double-
RT stranded DNA forms isolated from plants infected with tomato golden
RT mosaic virus.";
RL Nucleic Acids Res. 14:7967-7984(1986).
DR EMBL; X04485; CAA28171.1; -

```

DR InterPro: IPR000211; Gemini_BL.
DR Pfam: PF00845; Gemini_BLI; 1.
SQ SEQUENCE 55 AA; 6281 MW; 41963570D739EC96 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 12; Length 55;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 177 NVIESNRD 184
DB 13 NVIESNRD 20

RESULT 6
Q9HR77
ID Q9HR77 PRELIMINARY; PRT; 148 AA.
AC Q9HR77
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Vng0825c.
GN Vng0825c.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriia; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genome sequence of Halobacterium species NRC-1.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL: AE005024; AAG19281.1;
DR InterPro: IPR004435; MobB.
DR Pfam: PF03205; MobB; 1.
KW Complete proteome.
SQ SEQUENCE 148 AA; 15073 MW; F242336129C362FE CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 17; Length 148;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 TVVGPSSD 162
DB 5 TVVGPSSD 12

RESULT 7
Q9DGZ5
ID Q9DGZ5 PRELIMINARY; PRT; 154 AA.
AC Q9DGZ5
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE x protein.
GN x.
OS Hepatitis B virus.
OC Viruses; Retroviridae; Hepadnaviridae; Orthohepadnavirus.
OX NCBI_TaxID=10407;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=14118, AND 11141;
RA Hannoun C., Norder H., Lindh M.;
RT "An aberrant genotype revealed in recombinant hepatitis B virus
strains from Vietnam.";
SQ SEQUENCE 154 AA; 16563 MW; E6B37B6233218641B CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 12; Length 154;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 GDLPPSSS 55
DB 35 GDLPPSSS 42

RESULT 9
Q9ABT4
ID Q9ABT4 PRELIMINARY; PRT; 239 AA.
AC Q9ABT4
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein CC0136.
GN CC0136.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;

```

```

RL J. Gen. Virol. 81:2267-2272(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=14118, AND 11141;
RA Hannoun C., Norder H., Lindh M.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF241408; AAG17579.1;
DR EMBL: AF241407; AAG17572.1;
DR InterPro: IPR000236; TransactX.
DR Pfam: PF00739; X; 1.
SQ SEQUENCE 154 AA; 16516 MW; A13D1D6233218640D CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 12; Length 154;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 GDLPPSSS 55
DB 35 GDLPPSSS 42

RESULT 8
Q9E6T5
ID Q9E6T5 PRELIMINARY; PRT; 154 AA.
AC Q9E6T5
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE x protein.
GN x.
OS Hepatitis B virus.
OC Viruses; Retroviridae; Hepadnaviridae; Orthohepadnavirus.
OX NCBI_TaxID=10407;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=6871;
RX MEDLINE=20409088; PubMed=10950984;
RA Hannoun C., Norder H., Lindh M.;
RT "An aberrant genotype revealed in recombinant hepatitis B virus
strains from Vietnam.";
RL J. Gen. Virol. 81:2267-2272(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=6871;
RA Hannoun C., Norder H., Lindh M.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF241409; AAG17586.1;
DR InterPro: IPR000236; TransactX.
DR Pfam: PF00739; X; 1.
SQ SEQUENCE 154 AA; 16563 MW; E6B37B6233218641B CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 12; Length 154;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 GDLPPSSS 55
DB 35 GDLPPSSS 42

RESULT 9
Q9ABT4
ID Q9ABT4 PRELIMINARY; PRT; 239 AA.
AC Q9ABT4
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein CC0136.
GN CC0136.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;

```

```

RN  SEQUENCE FROM N.A.
RP  STRAIN-ATCC 19089 / CB15;
RX  MEDLINE-21173698; PubMed-11259647;
RA  Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA  Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA  Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA  DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA  Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA  Utterback T., Tran K., Wolf A., Vamathevan J., Esmolaeva M., White O.,
RA  Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT  "Complete genome sequence of Caulobacter crescentus.";
RL  Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR  EMBL; AE005688; AAK22123.1;
DR  TIGR; CC0136;
KW  Hypothetical protein; Complete proteome.
SQ  SEQUENCE 239 AA; 24704 MW; 1531A33ADACF9E31 CRC64;

Query Match 0.8%; Score 8; DB 16; Length 239;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 AEAPRRGR 25
Db 176 AEAPRRGR 183

RESULT 10
Q926W3
ID Q926W3 PRELIMINARY; PRT; 268 AA.
AC Q926W3;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical protein lin2927.
GN LIN2927.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-CLIP 11262 / SEROVAR 6A;
RX PubMed-11679669;
RA Baquero F., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Glaser P., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fshih H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL596174; CAC98152.1;
DR ListList; LIN02927;
DR InterPro; IPR000281; HTH_RpiR.
DR Pfam; PF01418; HTH_6; 1.
DR Pfam; PF01380; SIS; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 268 AA; 30656 MW; 7923711B8A818BE7 CRC64;

Query Match 0.8%; Score 8; DB 16; Length 268;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 792 LKNINKE 799
Db 32 LKNINKE 39

RESULT 12
Q97ZN3
ID Q97ZN3 PRELIMINARY; PRT; 281 AA.
AC Q97ZN3;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE C43H8.3 protein.
GN C43H8.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-BRISTOL N2;
RX MEDLINE-94150718; PubMed-7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Pavello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,

```

```

RESULT 11
Q8Y3P0
ID Q8Y3P0 PRELIMINARY; PRT; 268 AA.
AC Q8Y3P0;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical protein lmo2795.
GN LMO2795.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-EGD-E / SEROVAR 1/2A;
RX MEDLINE-21537279; PubMed-11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fshih H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL591984; CAD01008.1;
DR ListList; LMO02795;
DR InterPro; IPR000281; HTH_RpiR.
DR InterPro; IPR001347; SIS.
DR Pfam; PF01418; HTH_6; 1.
DR Pfam; PF01380; SIS; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 268 AA; 30643 MW; 7C26214E8FD1DAA2 CRC64;

Query Match 0.8%; Score 8; DB 16; Length 268;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 792 LKNINKE 799
Db 32 LKNINKE 39

RESULT 12
Q97ZN3
ID Q97ZN3 PRELIMINARY; PRT; 281 AA.
AC Q97ZN3;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE C43H8.3 protein.
GN C43H8.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-BRISTOL N2;
RX MEDLINE-94150718; PubMed-7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Pavello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,

```

RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,
RA Smalton N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaubin M., Vaughan K., Waterston R.,
RA Watson A., Weinscock L., Wilkinson-Sproat J., Wohlman P.,
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.*;
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Tin-Wollam A.M.;
RT "The sequence of C. elegans cosmid C43H8.*";
RL Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF098499; AAC67398.1; -;
DR InterPro: IPR001791; Laminin_G.
DR Pfam: PF00054; laminin_G; 1.
DR SMART: SM00282; LamG; 1.
SQ SEQUENCE 281 AA; 31708 MW; 5253CC040277343C CRC64;

Query Match 0.8%; Score 8; DB 5; Length 281;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 NITDFTVK 738
|||||||
Db 111 NITDFTVK 118

RESULT 13
Q9PR14 PRELIMINARY; PRT; 286 AA.
ID Q9PR14
AC Q9PR14
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ATP synthase gamma chain.
GN ATPG OR UUI30.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Ureaplasma.
OX NCBI_TaxID=134821;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SEROVAR 3;
RX MEDLINE=20500219; PubMed=11048724;
RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
RA Cassell G.H.;
RT "The complete sequence of the mucosal pathogen Ureaplasma
RT urealyticum.*";
RL Nature 407:757-762(2000).
DR EMBL; AE002114; AAF30536.1; -;
DR InterPro: IPR000131; ATPase_gamma.
DR Pfam: PF00231; ATP-synt; 1.
DR PRINTS; PR00126; ATPASEGAMMA.
DR TIGRFAMS; TIGR01146; ATPsyn_Flgamma; 1.
KW Complete proteome.
SQ SEQUENCE 286 AA; 32295 MW; C9AE278976AAE4F5 CRC64;

Query Match 0.8%; Score 8; DB 16; Length 286;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 739 EIVAGSD 746
|||||||
Db 279 EIVAGSD 286

RESULT 14

Q8V9I1 PRELIMINARY; PRT; 293 AA.
ID Q8V9I1
AC Q8V9I1
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Movement protein.
GN MP.
OS Squash yellow mottle virus.
OC Viruses; ssDNA viruses; Geminiviridae; Begomovirus.
OX NCBI_TaxID=91372;
RN [1]
RP SEQUENCE FROM N.A.
RA Ramirez P., Karkashian J., Zuniga C., Maxwell D.;
RT "Association of Squash Yellow Mottle Virus with cucurbits and papaya
RT in Costa Rica.*";
RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF440790; AAL33677.1; -;
DR InterPro: IPR000211; Geminin_BL.
DR Pfam: PF00845; Geminin_BL; 1.
SQ SEQUENCE 293 AA; 33078 MW; 8F99918F4FFE54AC CRC64;

Query Match 0.8%; Score 8; DB 12; Length 293;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 177 NYIESNRD 184
|||||||
Db 13 NYIESNRD 20

RESULT 15
Q50198 PRELIMINARY; PRT; 379 AA.
ID Q50198
AC Q50198
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE L222-ORF9.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97124199; PubMed=8969512;
RA Psihi H., De Rossi E., Salazar L., Cantoni R., Labo M., Riccardi G.,
RA Takiff H.E., Eiglmeyer K., Bergh S., Cole S.F.;
RT "Gene arrangement and organization in a approximately 76 kb fragment
RT encompassing the oric region of the chromosome of Mycobacterium
RT leprae.*";
RL Microbiology 142:0-0(0).
DR EMBL; L39923; AAB53129.1; -;
DR InterPro: IPR004268; MVIN_like.
DR Pfam: PF03023; MVIN; 1.
SQ SEQUENCE 379 AA; 40010 MW; FD23F9F20B4A74F5 CRC64;

Query Match 0.8%; Score 8; DB 2; Length 379;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 936 TAVLLVAL 943
|||||||
Db 243 TAVLLVAL 250

Search completed: May 12, 2003, 13:40:52
Job time : 35.4548 secs


```
QY 61 COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120
Db 61 COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120
QY 121 SKCGEGTSLGSGIKFDEWDELPAFNSIATFMDTVVGPDSRPPDCGNNSSWIPRGNYIE 180
Db 121 SKCGEGTSLGSGIKFDEWDELPAFNSIATFMDTVVGPDSRPPDCGNNSSWIPRGNYIE 180
QY 181 SNRDDCTVSLIYAVHLKSGYVFEFYQYVDNNIFFEFFFQNDQCOEMDITTDKWKVLTDN 240
Db 181 SNRDDCTVSLIYAVHLKSGYVFEFYQYVDNNIFFEFFFQNDQCOEMDITTDKWKVLTDN 240
QY 241 GEMGSHSVMLKSGTNLYWRTTGILMGSKAVPVLVKNITIEGVAYTSCFPCKPGTFNS 300
Db 241 GEMGSHSVMLKSGTNLYWRTTGILMGSKAVPVLVKNITIEGVAYTSCFPCKPGTFNS 300
QY 301 KPGSFNCQVCPRNTYSEKAKCIRCKDDSQFSGSECTERPPCTTKDYFQIHTPCDEBG 360
Db 301 KPGSFNCQVCPRNTYSEKAKCIRCKDDSQFSGSECTERPPCTTKDYFQIHTPCDEBG 360
QY 361 KTOIMYKWIPEKICREDLTDALRPPSGEKKDCPCPCNPGFYNNGSSCHPCPPGTFSDGT 420
Db 361 KTOIMYKWIPEKICREDLTDALRPPSGEKKDCPCPCNPGFYNNGSSCHPCPPGTFSDGT 420
QY 421 KECRPCPAGTEPALGFYKWNVLPGNMTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGS 480
Db 421 KECRPCPAGTEPALGFYKWNVLPGNMTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGS 480
QY 481 DNDYLILNLHIIPGFKPPTSMGTATGSELGRITFVFETLCSADCVLYFMVDINRKSTNVVE 540
Db 481 DNDYLILNLHIIPGFKPPTSMGTATGSELGRITFVFETLCSADCVLYFMVDINRKSTNVVE 540
QY 541 SWGGTKEKOAYTHIIFKNATFTTFAFORTNOGQDNRRFINDMVKIYSTTATNADGVAS 600
Db 541 SWGGTKEKOAYTHIIFKNATFTTFAFORTNOGQDNRRFINDMVKIYSTTATNADGVAS 600
QY 601 SCRACALGSEQSGSCVCPGPGHYIEKETNOCKECPDPTLYSIHQVYGKEACIPCGPGSK 660
Db 601 SCRACALGSEQSGSCVCPGPGHYIEKETNOCKECPDPTLYSIHQVYGKEACIPCGPGSK 660
QY 661 NNQDHSVCYSDCFFYHEKENQILHYDFSNLSSVGLMNGPSTSKGTYEHFFNLSLCH 720
Db 661 NNQDHSVCYSDCFFYHEKENQILHYDFSNLSSVGLMNGPSTSKGTYEHFFNLSLCH 720
QY 721 EGKKNALCTNNITDFTVKEIVAGSDDTYTLVGAFCVQSTIIPSEKGFRAALSSQSIILA 780
Db 721 EGKKNALCTNNITDFTVKEIVAGSDDTYTLVGAFCVQSTIIPSEKGFRAALSSQSIILA 780
QY 781 DTFIGVTVETTLKNNIKEDMEPVPTSQIPDVHFFYKKSSTATTSCLNGRSTAVKMRCPNT 840
Db 781 DTFIGVTVETTLKNNIKEDMEPVPTSQIPDVHFFYKKSSTATTSCLNGRSTAVKMRCPNT 840
QY 841 KSGAGVISVPSKCPAGTCDCGCTFYFLWESAECPLCTEHDHFIEGACKRGFOETLYVWN 900
Db 841 KSGAGVISVPSKCPAGTCDCGCTFYFLWESAECPLCTEHDHFIEGACKRGFOETLYVWN 900
QY 901 EPKWCITGSLPEKKLATCETVDFWLKVAGVGAFTAVLLVALTCYFMKKNQKLEYKYSK 960
Db 901 EPKWCITGSLPEKKLATCETVDFWLKVAGVGAFTAVLLVALTCYFMKKNQKLEYKYSK 960
QY 961 LYMTTNSKECELPAAADSCAIMEGEONEEVVYSNQSLGKLSLATKEKEDHFPESVOLK 1020
Db 961 LYMTTNSKECELPAAADSCAIMEGEONEEVVYSNQSLGKLSLATKEKEDHFPESVOLK 1020
QY 1021 TSRSNPNI 1027
Db 1021 TSRSNPNI 1027
```

RESULT 2
US-10-140-164-2
: Sequence 2, Application US/10140164
: Publication No. US20030072736A1

```
: GENERAL INFORMATION:  
: APPLICANT: Baker et al.  
: TITLE OF INVENTION: Human Tumor Necrosis Factor Receptor Tr16  
: FILE REFERENCE: PF514C1  
: CURRENT APPLICATION NUMBER: US/10/140,164  
: CURRENT FILING DATE: 2002-03-08  
: PRIOR APPLICATION NUMBER: 09/637,856  
: PRIOR FILING DATE: 2000-08-10  
: PRIOR APPLICATION NUMBER: 60/148,348  
: PRIOR FILING DATE: 1999-08-12  
: PRIOR APPLICATION NUMBER: 60/148,683  
: PRIOR FILING DATE: 1999-08-13  
: PRIOR APPLICATION NUMBER: 60/148,870  
: PRIOR FILING DATE: 1999-08-13  
: PRIOR APPLICATION NUMBER: 60/148,758  
: PRIOR FILING DATE: 1999-08-16  
: PRIOR APPLICATION NUMBER: 60/149,181  
: PRIOR FILING DATE: 1999-08-17  
: PRIOR APPLICATION NUMBER: 60/149,453  
: PRIOR FILING DATE: 1999-08-18  
: PRIOR APPLICATION NUMBER: 60/149,498  
: PRIOR FILING DATE: 1999-08-19  
: NUMBER OF SEQ ID NOS: 76  
: SOFTWARE: PatentIn Ver. 2.1  
: SEQ ID NO 2  
: LENGTH: 963  
: TYPE: PRT  
: ORGANISM: Homo sapiens  
US-10-140-164-2
```

Query Match 92.8%; Score 953; DB 9; Length 963;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 953; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MLFRARGPVRGWRGAPAEAPRRGRSPWPSPAWICWALAGCAAAGDLPSSSRPLPP 60
Db 1 MLFRARGPVRGWRGAPAEAPRRGRSPWPSPAWICWALAGCAAAGDLPSSSRPLPP 60
QY 61 COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120
Db 61 COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120
QY 121 SKCGEGTSLGSGIKFDEWDELPAFNSIATFMDTVVGPDSRPPDCGNNSSWIPRGNYIE 180
Db 121 SKCGEGTSLGSGIKFDEWDELPAFNSIATFMDTVVGPDSRPPDCGNNSSWIPRGNYIE 180
QY 181 SNRDDCTVSLIYAVHLKSGYVFEFYQYVDNNIFFEFFFQNDQCOEMDITTDKWKVLTDN 240
Db 181 SNRDDCTVSLIYAVHLKSGYVFEFYQYVDNNIFFEFFFQNDQCOEMDITTDKWKVLTDN 240
QY 241 GEMGSHSVMLKSGTNLYWRTTGILMGSKAVPVLVKNITIEGVAYTSCFPCKPGTFNS 300
Db 241 GEMGSHSVMLKSGTNLYWRTTGILMGSKAVPVLVKNITIEGVAYTSCFPCKPGTFNS 300
QY 301 KPGSFNCQVCPRNTYSEKAKCIRCKDDSQFSGSECTERPPCTTKDYFQIHTPCDEBG 360
Db 301 KPGSFNCQVCPRNTYSEKAKCIRCKDDSQFSGSECTERPPCTTKDYFQIHTPCDEBG 360
QY 361 KTOIMYKWIPEKICREDLTDALRPPSGEKKDCPCPCNPGFYNNGSSCHPCPPGTFSDGT 420
Db 361 KTOIMYKWIPEKICREDLTDALRPPSGEKKDCPCPCNPGFYNNGSSCHPCPPGTFSDGT 420
QY 421 KECRPCPAGTEPALGFYKWNVLPGNMTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGS 480
Db 421 KECRPCPAGTEPALGFYKWNVLPGNMTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGS 480
QY 481 DNDYLILNLHIIPGFKPPTSMGTATGSELGRITFVFETLCSADCVLYFMVDINRKSTNVVE 540
Db 481 DNDYLILNLHIIPGFKPPTSMGTATGSELGRITFVFETLCSADCVLYFMVDINRKSTNVVE 540
QY 541 SWGGTKEKOAYTHIIFKNATFTTFAFORTNOGQDNRRFINDMVKIYSTTATNADGVAS 600
Db 541 SWGGTKEKOAYTHIIFKNATFTTFAFORTNOGQDNRRFINDMVKIYSTTATNADGVAS 600
```

QY 601 SCRACALGSEQSGSCVPCPPGHYIEKETNOCKECPDPTLYLSIHQYVYKEACIPCGPGSK 660
Db 601 SCRACALGSEQSGSCVPCPPGHYIEKETNOCKECPDPTLYLSIHQYVYKEACIPCGPGSK 660
QY 661 NNQDHSVYSDCFFVHEKENQILHDFNSLSSVGLMNGPSTSGTKYFHFNFISLGH 720
Db 661 NNQDHSVYSDCFFVHEKENQILHDFNSLSSVGLMNGPSTSGTKYFHFNFISLGH 720
QY 721 EGKMACTNNITDFTVKIIVAGSDDYTNLVGAFVCQSTIIPSESKGPRAAALSSQSIIIA 780
Db 721 EGKMACTNNITDFTVKIIVAGSDDYTNLVGAFVCQSTIIPSESKGPRAAALSSQSIIIA 780
QY 781 DTFIGVYVETTLKINIKEDMPVPTSQIPDVHFFYKSTATTSCINGRSTAVKMRNPT 840
Db 781 DTFIGVYVETTLKINIKEDMPVPTSQIPDVHFFYKSTATTSCINGRSTAVKMRNPT 840
QY 841 KSGAGVISVPSKCPAGTCDGCTFYFLWESAECPLCTEHDHFEIEGACKRGFOETLYVWN 900
Db 841 KSGAGVISVPSKCPAGTCDGCTFYFLWESAECPLCTEHDHFEIEGACKRGFOETLYVWN 900
QY 901 EPKWCITKISLPEKKLATCETVDFWLKVGAGVGAFTAVLLVALTCTYFVKKNOK 953
Db 901 EPKWCITKISLPEKKLATCETVDFWLKVGAGVGAFTAVLLVALTCTYFVKKNOK 953

RESULT 3
US-10-002-050-20
; Sequence 20, Application US/10002050
; Publication No. US20030032095A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corine
; APPLICANT: Yang, Meijia
; APPLICANT: Boldog, Ferenc
; APPLICANT: Herrmann, John
; TITLE OF INVENTION: No: US20030032095A1el 'Nucleic Acid Sequences Encoding Human Sena
; FILE REFERENCE: 15966-554 Cura-S14
; CURRENT APPLICATION NUMBER: US/10/002,050
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 09/604,286
; PRIOR FILING DATE: 2000-06-22
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-002-050-20

Query Match 26.4%; Score 271; DB 9; Length 464;
Best Local Similarity 99.6%; Pred. No. 1.2e-260;
Matches 451; Conservative 0; Mismatches 0; Indels 2; Gaps 1;
QY 115 MKNQVCSKCGEGTSLGSGIKFDEWDELPAFNSNIATFMDTVVGPDSRPGDGNSSWIP 174
Db 1 MKNQVCSKCGEGTSLGSGIKFDEWDELPAFNSNIATFMDTVVGPDSRPGDGNSSWIP 60
QY 175 RGNYESNRDDCTVSLIYAVHLKKSQYVFEYQYVDNNIFFEFFFIONDQCEMDTTTDKW 234
Db 61 RGNYESNRDDCTVSLIYAVHLKKSQYVFEYQYVDNNIFFEFFFIONDQCEMDTTTDKW 120
QY 235 VKLTDNGEGSHVLMKSGTNILYWRITGILMGSKAVKPVLVKNITIEGVAYTSECFPCK 294
Db 121 VKLTDNGEGSHVLMKSGTNILYWRITGILMGSKAVKPVLVKNITIEGVAYTSECFPCK 180
QY 295 PGTFSNKPGSFNCQVCPRTYSEKGAKEICRCKDDSOFS--GSSECTERPPCTTKDYFOI 352
Db 181 PGTFSNKPGSFNCQVCPRTYSEKGAKEICRCKDDSOFS--GSSECTERPPCTTKDYFOI 240
QY 353 HTPCDEEGTQIMYKWIIEPKICREDLTDALRPPSGEKKDCPCNPGFYNNNGSSSCHPCP 412
Db 241 HTPCDEEGTQIMYKWIIEPKICREDLTDALRPPSGEKKDCPCNPGFYNNNGSSSCHPCP 300
QY 413 PGTFSDGTKECRPCPAGTEPALGFYKWNVLPGNNKTSFCFNVGNSKCDGMNGWEVAGDH 472
Db 301 PGTFSDGTKECRPCPAGTEPALGFYKWNVLPGNNKTSFCFNVGNSKCDGMNGWEVAGDH 360
QY 473 IQSGAGGSDNDYLILNLHIPGPKPPTSMTGATGSELGRITTFVETILCSADCVLYFNVWDIN 532
Db 240 IQSGAGGSDNDYLILNLHIPGPKPPTSMTGATGSELGRITTFVETILCSADCVLYFNVWDIN 420

QY 353 HTPCDEEGTQIMYKWIIEPKICREDLTDALRPPSGEKKDCPCNPGFYNNNGSSSCHPCP 412
Db 241 HTPCDEEGTQIMYKWIIEPKICREDLTDALRPPSGEKKDCPCNPGFYNNNGSSSCHPCP 300
QY 413 PGTFSDGTKECRPCPAGTEPALGFYKWNVLPGNNKTSFCFNVGNSKCDGMNGWEVAGDH 472
Db 301 PGTFSDGTKECRPCPAGTEPALGFYKWNVLPGNNKTSFCFNVGNSKCDGMNGWEVAGDH 360
QY 473 IQSGAGGSDNDYLILNLHIPGPKPPTSMTGATGSELGRITTFVETILCSADCVLYFNVWDIN 532
Db 361 IQSGAGGSDNDYLILNLHIPGPKPPTSMTGATGSELGRITTFVETILCSADCVLYFNVWDIN 420
QY 533 RKSTNVVESWGCTKEQAYTHIIFKNATFTFTW 565
Db 421 RKSTNVVESWGCTKEQAYTHIIFKNATFTFTW 453

RESULT 4
US-10-002-304-20
; Sequence 20, Application US/10002304
; Publication No. US20030036185A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corine
; APPLICANT: Yang, Meijia
; APPLICANT: Boldog, Ferenc
; APPLICANT: Herrmann, John
; TITLE OF INVENTION: Polynucleotides and polypeptides encoded thereby
; FILE REFERENCE: 15966-554 Cura-S8
; CURRENT APPLICATION NUMBER: US/10/002,304
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 09/604,286
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/140,584
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-002-304-20

Query Match 26.4%; Score 271; DB 9; Length 464;
Best Local Similarity 99.6%; Pred. No. 1.2e-260;
Matches 451; Conservative 0; Mismatches 0; Indels 2; Gaps 1;
QY 115 MKNQVCSKCGEGTSLGSGIKFDEWDELPAFNSNIATFMDTVVGPDSRPGDGNSSWIP 174
Db 1 MKNQVCSKCGEGTSLGSGIKFDEWDELPAFNSNIATFMDTVVGPDSRPGDGNSSWIP 60
QY 175 RGNYESNRDDCTVSLIYAVHLKKSQYVFEYQYVDNNIFFEFFFIONDQCEMDTTTDKW 234
Db 61 RGNYESNRDDCTVSLIYAVHLKKSQYVFEYQYVDNNIFFEFFFIONDQCEMDTTTDKW 120
QY 235 VKLTDNGEGSHVLMKSGTNILYWRITGILMGSKAVKPVLVKNITIEGVAYTSECFPCK 294
Db 121 VKLTDNGEGSHVLMKSGTNILYWRITGILMGSKAVKPVLVKNITIEGVAYTSECFPCK 180
QY 295 PGTFSNKPGSFNCQVCPRTYSEKGAKEICRCKDDSOFS--GSSECTERPPCTTKDYFOI 352
Db 181 PGTFSNKPGSFNCQVCPRTYSEKGAKEICRCKDDSOFS--GSSECTERPPCTTKDYFOI 240
QY 353 HTPCDEEGTQIMYKWIIEPKICREDLTDALRPPSGEKKDCPCNPGFYNNNGSSSCHPCP 412
Db 241 HTPCDEEGTQIMYKWIIEPKICREDLTDALRPPSGEKKDCPCNPGFYNNNGSSSCHPCP 300
QY 413 PGTFSDGTKECRPCPAGTEPALGFYKWNVLPGNNKTSFCFNVGNSKCDGMNGWEVAGDH 472
Db 301 PGTFSDGTKECRPCPAGTEPALGFYKWNVLPGNNKTSFCFNVGNSKCDGMNGWEVAGDH 360
QY 473 IQSGAGGSDNDYLILNLHIPGPKPPTSMTGATGSELGRITTFVETILCSADCVLYFNVWDIN 532
Db 420 IQSGAGGSDNDYLILNLHIPGPKPPTSMTGATGSELGRITTFVETILCSADCVLYFNVWDIN 420

|||||
Db 361 IOSGAGSDNDYLILNLHIFGPKPTSMTGATGSELGRITFVFETLCSADCVLYFMVDIN 420
QY 533 RKSTNVVSWGGTKEKQAYTHIIFKNATFTTW 565
Db 421 RKSTNVVSWGGTKEKQAYTHIIFKNATFTTW 453

RESULT 5

US-10-003-152-20
; Sequence 20, Application US/10003152
; Patent No. US20020151494A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corine
; APPLICANT: Yang, Meljia
; APPLICANT: Boldog, Ferenc
; APPLICANT: Herrmann, John
; TITLE OF INVENTION: No. US20020151494A1 Amino Acid Sequences for Human Semaphorin-1
; FILE REFERENCE: 15966-554 Cura-54 CON-S12
; CURRENT APPLICATION NUMBER: US/10/003,152
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 09/604,286
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/140,584
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-003-152-20

Query Match 26.4%; Score 271; DB 12; Length 464;
Best Local Similarity 99.6%; Pred. No. 1.2e-260;
Matches 451; Conservative 0; Mismatches 0; Indels 2; Gaps 1;
QY 115 MKNQVCSKCGEGTSLGSGIKKIDFEMDELPAFNSNIATFMDTVVGPDSRPGDCGNNSSWIP 174
Db 1 MKNQVCSKCGEGTSLGSGIKKIDFEMDELPAFNSNIATFMDTVVGPDSRPGDCGNNSSWIP 60
QY 175 RGNVIESNRDDCTVSLIYAVHLKKSQYVFYQYVDNNITFFEFFIIONDQCQEMDTTDDKW 234
Db 61 RGNVIESNRDDCTVSLIYAVHLKKSQYVFYQYVDNNITFFEFFIIONDQCQEMDTTDDKW 120
QY 235 VKLTDNGENSHVMUKSGTNNILYWTGILMGSKAVKPVLYKNITIEGVAYTSECFPCPK 294
Db 121 VKLTDNGENSHVMUKSGTNNILYWTGILMGSKAVKPVLYKNITIEGVAYTSECFPCPK 180
QY 295 PGTFSNKPQSFNCQVCPRTYSEKGAKEICRCKDDSQFS--GSSECTERPCTTKDYFOI 352
Db 181 PGTFSNKPQSFNCQVCPRTYSEKGAKEICRCKDDSQFSSESECTERPCTTKDYFOI 240
QY 353 HTPCDEBEGTKQIMYKWIPIKREDLTDAILPPSGEKKDCPCPNPGFYNNSSSCHPCP 412
Db 241 HTPCDEBEGTKQIMYKWIPIKREDLTDAILPPSGEKKDCPCPNPGFYNNSSSCHPCP 300
QY 413 PGTFSDGTEKPCPCPAGTEPALGFYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDH 472
Db 301 PGTFSDGTEKPCPCPAGTEPALGFYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDH 360
QY 473 IOSGAGSDNDYLILNLHIFGPKPTSMTGATGSELGRITFVFETLCSADCVLYFMVDIN 532
Db 361 IOSGAGSDNDYLILNLHIFGPKPTSMTGATGSELGRITFVFETLCSADCVLYFMVDIN 420
QY 533 RKSTNVVSWGGTKEKQAYTHIIFKNATFTTW 565
Db 421 RKSTNVVSWGGTKEKQAYTHIIFKNATFTTW 453

RESULT 6

US-10-002-050-10
; Sequence 10, Application US/10002050
; Publication No. US20030032095A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corine
; APPLICANT: Yang, Meljia
; APPLICANT: Boldog, Ferenc
; APPLICANT: Herrmann, John
; TITLE OF INVENTION: No. US20030032095A1el Nucleic Acid Sequences Encoding Human Se
; FILE REFERENCE: 15966-554 Cura-54 CON-S14
; CURRENT APPLICATION NUMBER: US/10/002,050
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 09/604,286
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/140,584
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-002-050-10

Query Match 22.6%; Score 232; DB 9; Length 411;
Best Local Similarity 100.0%; Pred. No. 6.4e-222;
Matches 232; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 GSSECTERPCTTKDYFOIHTPCDEBEGTKQIMYKWIPIKREDLTDAILRPPSGEKKDC 393
Db 169 GSSECTERPCTTKDYFOIHTPCDEBEGTKQIMYKWIPIKREDLTDAILRPPSGEKKDC 228
QY 394 PCNPGFYNNSSSCHPCPPTGTFSDGTEKPCPCPAGTEPALGFYKWNVLPGNMKTSCF 453
Db 229 PCNPGFYNNSSSCHPCPPTGTFSDGTEKPCPCPAGTEPALGFYKWNVLPGNMKTSCF 288
QY 454 NVGNSKCDGMNGWEVAGDHITQSGAGSDNDYLILNLHIFGPKPTSMTGATGSELGRITF 513
Db 289 NVGNSKCDGMNGWEVAGDHITQSGAGSDNDYLILNLHIFGPKPTSMTGATGSELGRITF 348
QY 514 VFETLCSADCVLYFMVDINRKSTNVVSWGGTKEKQAYTHIIFKNATFTTW 565
Db 349 VFETLCSADCVLYFMVDINRKSTNVVSWGGTKEKQAYTHIIFKNATFTTW 400

RESULT 7

US-10-002-304-10
; Sequence 10, Application US/10002304
; Publication No. US20030036185A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corine
; APPLICANT: Yang, Meljia
; APPLICANT: Boldog, Ferenc
; APPLICANT: Herrmann, John
; TITLE OF INVENTION: Polynucleotides and polypeptides encoded thereby
; FILE REFERENCE: 15966-554 Cura-54 CON-S8
; CURRENT APPLICATION NUMBER: US/10/002,304
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 09/604,286
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/140,584
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-002-304-10

```
Query Match      22.6%; Score 232; DB 9; Length 411;
Best Local Similarity 100.0%; Pred. No. 6.4e-222;
Matches 232; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 GSSECTERPPCTTKDYFOIHTPCDEEGKTQIMYKWIPEKICREDLTDAILRPPSGEKKDC 393
|||||
Db 169 GSSECTERPPCTTKDYFOIHTPCDEEGKTQIMYKWIPEKICREDLTDAILRPPSGEKKDC 228
|||||

QY 394 PCNPGFYNNSSSCHPCPPGTFSDGTEKCRPCPCAGTEPAGLGFYKWNVLPGNMKTSCF 453
|||||
Db 229 PCNPGFYNNSSSCHPCPPGTFSDGTEKCRPCPCAGTEPAGLGFYKWNVLPGNMKTSCF 288
|||||

QY 454 NVGNSKCDGMNGWEVAGDHIOGAGGSDNDYLILNLHIPGFKPPTSMTGATGSELGRITF 513
|||||
Db 289 NVGNSKCDGMNGWEVAGDHIOGAGGSDNDYLILNLHIPGFKPPTSMTGATGSELGRITF 348
|||||

QY 514 VFETLCSADCVLYFMVDINRKNSTNVVSGGTKEKOAYTHIIFKNATFTFW 565
|||||
Db 349 VFETLCSADCVLYFMVDINRKNSTNVVSGGTKEKOAYTHIIFKNATFTFW 400
|||||

RESULT 8
US-10-003-152-10
; Sequence 10, Application US/10003152
; Patent No. US20020151494A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corine
; APPLICANT: Yang, Meijia
; APPLICANT: Boldog, Ferenc
; APPLICANT: Herrmann, John
; TITLE OF INVENTION: No. US20020151494A1el Amino Acid Sequences for Human Semaphorin-1
; FILE REFERENCE: 15966-554 Cura-54 CON-512
; CURRENT APPLICATION NUMBER: US/10/003,152
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 09/604,286
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/140,584
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-003-152-10

Query Match      22.6%; Score 232; DB 12; Length 411;
Best Local Similarity 100.0%; Pred. No. 6.4e-222;
Matches 232; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 GSSECTERPPCTTKDYFOIHTPCDEEGKTQIMYKWIPEKICREDLTDAILRPPSGEKKDC 393
|||||
Db 169 GSSECTERPPCTTKDYFOIHTPCDEEGKTQIMYKWIPEKICREDLTDAILRPPSGEKKDC 228
|||||

QY 394 PCNPGFYNNSSSCHPCPPGTFSDGTEKCRPCPCAGTEPAGLGFYKWNVLPGNMKTSCF 453
|||||
Db 229 PCNPGFYNNSSSCHPCPPGTFSDGTEKCRPCPCAGTEPAGLGFYKWNVLPGNMKTSCF 288
|||||

QY 454 NVGNSKCDGMNGWEVAGDHIOGAGGSDNDYLILNLHIPGFKPPTSMTGATGSELGRITF 513
|||||
Db 289 NVGNSKCDGMNGWEVAGDHIOGAGGSDNDYLILNLHIPGFKPPTSMTGATGSELGRITF 348
|||||

QY 514 VFETLCSADCVLYFMVDINRKNSTNVVSGGTKEKOAYTHIIFKNATFTFW 565
|||||
Db 349 VFETLCSADCVLYFMVDINRKNSTNVVSGGTKEKOAYTHIIFKNATFTFW 400
|||||

RESULT 9
US-09-864-761-39769
; Sequence 39769, Application US/09864761
```

```
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecmics-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 39769
; LENGTH: 81
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC002081.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
; OTHER INFORMATION: EST_HUMAN HIT: AW954806.1, EVALUE 5.00e-45
; OTHER INFORMATION: SWISSPROT HIT: P01267, EVALUE 3.00e-03
US-09-864-761-39769

Query Match      7.9%; Score 81; DB 10; Length 81;
Best Local Similarity 100.0%; Pred. No. 1.9e-72;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 578 RFINDMVKIYSITATNAVGVASSCRACALGSEQSGSSCVPCPPGHIYEKETNCKECP 637
|||||
Db 1 RFINDMVKIYSITATNAVGVASSCRACALGSEQSGSSCVPCPPGHIYEKETNCKECP 60
|||||
```

```
QY 638 DTYLSIHQVYKREACTPCPG 658
      |||||||
Db 61 DTYLSIHQVYKREACTPCPG 81

RESULT 10
US-10-140-164-36
; Sequence 36, Application US/10140164
; Publication No. US20030072736A1
; GENERAL INFORMATION:
; APPLICANT: Baker et al.
; TITLE OF INVENTION: Human Tumor Necrosis Factor Receptor TR16
; FILE REFERENCE: PF514C1
; CURRENT APPLICATION NUMBER: US/10/140,164
; CURRENT FILING DATE: 2002-05-08
; PRIOR APPLICATION NUMBER: 09/637,856
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/148,348
; PRIOR FILING DATE: 1999-08-12
; PRIOR APPLICATION NUMBER: 60/148,683
; PRIOR FILING DATE: 1999-08-13
; PRIOR APPLICATION NUMBER: 60/148,870
; PRIOR FILING DATE: 1999-08-13
; PRIOR APPLICATION NUMBER: 60/148,758
; PRIOR FILING DATE: 1999-08-16
; PRIOR APPLICATION NUMBER: 60/149,181
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: 60/149,453
; PRIOR FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: 60/149,498
; PRIOR FILING DATE: 1999-08-19
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 36
; LENGTH: 78
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-140-164-36
Query Match
Best Local Similarity 100.0%; Pred. No. 1.7e-69;
Matches 78; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 950 KNOKLEYKSKLVMTNTNSKECELPAAADSCAIMEGEDNEEEVYVSNKQSLGKLSLATKE 1009
      |||||||
Db 1 KNOKLEYKSKLVMTNTNSKECELPAAADSCAIMEGEDNEEEVYVSNKQSLGKLSLATKE 60

QY 1010 KEDHFESVOLKTSRSPNI 1027
      |||||||
Db 61 KEDHFESVOLKTSRSPNI 78

RESULT 12
US-10-140-164-32
; Sequence 32, Application US/10140164
; Publication No. US20030072736A1
; GENERAL INFORMATION:
; APPLICANT: Baker et al.
; TITLE OF INVENTION: Human Tumor Necrosis Factor Receptor TR16
; FILE REFERENCE: PF514C1
; CURRENT APPLICATION NUMBER: US/10/140,164
; CURRENT FILING DATE: 2002-05-08
; PRIOR APPLICATION NUMBER: 09/637,856
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/148,348
; PRIOR FILING DATE: 1999-08-12
; PRIOR APPLICATION NUMBER: 60/148,683
; PRIOR FILING DATE: 1999-08-13
; PRIOR APPLICATION NUMBER: 60/148,870
; PRIOR FILING DATE: 1999-08-13
; PRIOR APPLICATION NUMBER: 60/148,758
; PRIOR FILING DATE: 1999-08-16
; PRIOR APPLICATION NUMBER: 60/149,181
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: 60/149,453
; PRIOR FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: 60/149,498
; PRIOR FILING DATE: 1999-08-19
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 32
; LENGTH: 74
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-140-164-32
Query Match
Best Local Similarity 100.0%; Pred. No. 1.6e-65;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 338 CTERPCTTKDYFQIHTPCDEEGKTOIMYKWIPEKTCREDLTDAIRLPPSGEKKDCPPCN 397
      |||||||
Db 1 CTERPCTTKDYFQIHTPCDEEGKTOIMYKWIPEKTCREDLTDAIRLPPSGEKKDCPPCN 60

QY 398 PGFYNNSSSSCHPC 411
      |||||||
Db 61 PGFYNNSSSSCHPC 74
```

RESULT 13
US-10-140-164-61
; Sequence 61, Application US/10140164
; Publication No. US20030072736A1
; GENERAL INFORMATION:
; APPLICANT: Baker et al.
; TITLE OF INVENTION: Human Tumor Necrosis Factor Receptor TR16
; FILE REFERENCE: PF514C1
; CURRENT APPLICATION NUMBER: US/10/140,164
; CURRENT FILING DATE: 2002-05-08
; PRIOR APPLICATION NUMBER: 09/637,856
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/148,348
; PRIOR FILING DATE: 1999-08-12
; PRIOR APPLICATION NUMBER: 60/148,683
; PRIOR FILING DATE: 1999-08-13
; PRIOR APPLICATION NUMBER: 60/148,870
; PRIOR FILING DATE: 1999-08-13
; PRIOR APPLICATION NUMBER: 60/148,758
; PRIOR FILING DATE: 1999-08-16
; PRIOR APPLICATION NUMBER: 60/149,181
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: 60/149,453
; PRIOR FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: 60/149,498
; PRIOR FILING DATE: 1999-08-19
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 61
; LENGTH: 74
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-140-164-61

Query Match 7.2%; Score 74; DB 9; Length 74;
Best Local Similarity 100.0%; Pred. No. 1.6e-65;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 338 CTERPCTTKDYFQHTPCDEEGKQIMYKWIETPKICREDLDTAIRLPPSGEKKDCPPCN 397
|||||
Db 1 CTERPCTTKDYFQHTPCDEEGKQIMYKWIETPKICREDLDTAIRLPPSGEKKDCPPCN 60
|||||
QY 398 PGFYNGSSSCHPC 411
|||||
Db 61 PGFYNGSSSCHPC 74
|||||

RESULT 14
US-09-864-761-47095
; Sequence 47095, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemica-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/508,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 47095
; LENGTH: 64
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC002081.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.55
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.52
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.49
; OTHER INFORMATION: SWISSPROT HIT: Q00019, EVALUE 8.90e-01
US-09-864-761-47095

Query Match 6.2%; Score 64; DB 10; Length 64;
Best Local Similarity 100.0%; Pred. No. 1.2e-55;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 105 FSCASGEYLEMKNOVCSCGEGTYSGLSGIKFDEWDELPAFSGNSIATFMDTVVGPDSR 164
|||||
Db 1 FSCASGEYLEMKNOVCSCGEGTYSGLSGIKFDEWDELPAFSGNSIATFMDTVVGPDSR 60
|||||
QY 165 DGCN 168
|||||
Db 61 DGCN 64
|||||

RESULT 15
US-09-864-761-39057
; Sequence 39057, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemica-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6

```
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/006666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 39057
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC002081.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6
; OTHER INFORMATION: EST_HUMAN HIT: AW966212.1, EVALUE 2.00e-10
; OTHER INFORMATION: SWISSPROT HIT: P21849, EVALUE 5.00e-03
US-09-864-761-39057

Query Match          5.8%; Score 60; DB 10; Length 60;
Best Local Similarity 100.0%; Pred. No. 1.1e-51;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 TQIMYKWIPEKICREDLTAIRLPPSGEKKDCPPCNPGEYNNNGSSSCHPCPPCTFSDGTK 421
Db 1 TQIMYKWIPEKICREDLTAIRLPPSGEKKDCPPCNPGEYNNNGSSSCHPCPPCTFSDGTK 60

Search completed: May 12, 2003, 13:49:26
Job time : 23.1271 secs
```

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run On: May 12, 2003, 13:23:08 ; Search time 36.1256 seconds
(without alignments)
3788.125 Million cell updates/sec

Title: US-10-073-333A-4

Perfect score: 1027

Sequence: 1 MLFRAGPVRGNGRPAEA.....KEKEDHFSVOLKTRSPNI 1027

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : _A_Geneseq_101002.*

1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1027	100.0	1027	22 AAB70256	TR16-long receptor
2	953	92.8	963	22 AAB70255	TR16-short recepto
3	271	26.4	464	22 AAB48377	Human SEC10 protei
4	232	22.6	411	22 AAB48372	Human SEC5 protein
5	81	7.9	81	22 AAB39918	Peptide #7424 enco
6	81	7.9	81	22 AAB24471	Protein #6470 enco
7	81	7.9	81	22 AAB60663	Human brain expres
8	81	7.9	81	22 AAM73335	Human bone marrow
9	81	7.9	81	22 AAM33535	Peptide #7572 enco
10	81	7.9	81	23 ABG43186	Human peptide enco

11	78	7.6	78	22	AAB70285	Peptide #29. Unid
12	74	7.2	74	22	AAB70281	Peptide #25. Unid
13	64	6.2	64	22	AAM72925	Human bone marrow
14	64	6.2	64	22	ABG42760	Human peptide enco
15	60	5.8	60	22	ABG38686	Human peptide enco
16	60	5.8	60	22	ABB23759	Protein #5758 enco
17	60	5.8	60	22	AAM59318	Human brain expres
18	60	5.8	60	22	AAM71867	Human bone marrow
19	60	5.8	60	22	AAM32149	Peptide #6186 enco
20	60	5.8	60	22	ABG41680	Human peptide enco
21	50	4.9	50	22	ABB39681	Peptide #7187 enco
22	50	4.9	50	22	ABB24346	Protein #6345 enco
23	50	4.9	50	22	AAM60397	Human brain expres
24	50	4.9	50	22	AAM73033	Human bone marrow
25	50	4.9	50	22	AAM19811	Peptide #6245 enco
26	50	4.9	50	22	AAM33257	Peptide #7294 enco
27	50	4.9	50	22	ABG42877	Human peptide enco
28	37	3.6	71	22	AAU21345	Human novel foetal
29	21	2.0	105	21	AAB26180	Human CASB619 prot
30	21	2.0	495	20	AAY59972	Human endometrium
31	21	2.0	750	22	AAB35328	Human TR13 recepto
32	21	2.0	1001	22	AAB35333	Human CASB619 prot
33	21	2.0	1013	21	AAB26179	Human TR13 recepto
34	21	2.0	1013	22	AAU12190	Human PRO4985 poly
35	21	2.0	1013	22	AAB83845	Amino acid sequenc
36	15	1.5	372	22	AAB85768	Human seven-transm
37	15	1.5	383	22	AAB83853	Amino acid sequenc
38	15	1.5	870	22	AAB83851	Amino acid sequenc
39	15	1.5	911	22	AAB83850	Amino acid sequenc
40	11	1.1	52	22	AAB83848	Peptide fragment o
41	11	1.1	209	22	AAB83852	Amino acid sequenc
42	9	0.9	9	21	AAB26190	Human CASB619 prot
43	9	0.9	212	22	ABG04843	Novel human diagno
44	9	0.9	708	22	AAG91270	C. glutamicum prote
45	9	0.9	1487	19	AAW76719	Thermococcus 9N2 b

ALIGNMENTS

RESULT 1
AAB70256
ID AAB70256 standard; protein; 1027 AA.
XX
AC AAB70256;
XX
DT 10-MAY-2001 (first entry)
XX
DE TR16-long receptor protein.
XX
KW TR16 receptor; tumour necrosis factor receptor superfamily;
XX apoptosis; inflammatory; cancer; immune; neurodegenerative.
OS Unidentified.
XX
PN WO200112671-A1.
XX
PD 22-FEB-2001.
XX
PF 10-AUG-2000; 2000WO-US21885.
XX
PR 12-AUG-1999; 99US-0148348.
PR 13-AUG-1999; 99US-0148683.
PR 13-AUG-1999; 99US-0148870.
PR 16-AUG-1999; 99US-0148758.
PR 17-AUG-1999; 99US-0149181.
PR 18-AUG-1999; 99US-0149453.
PR 19-AUG-1999; 99US-0149498.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Young PE, Baker KP;
XX

DR WPI: 2001-138754/14.
XX New nucleic acid molecule encoding a TR16 tumor necrosis factor
PT receptor polypeptide, useful for the diagnosis and treatment of cancer,
PT autoimmune disorders and cardiovascular diseases -
XX
PS Disclosure: Fig 4; 286pp; English.
XX The present invention relates to a TR16 receptor (tumour necrosis
CC factor receptor superfamily). The invention is useful treating
CC diseases and disorders associated with the inhibited or increased
CC apoptosis. In particular inflammatory diseases, cancers, immune and
CC neurodegenerative disorders may be treated.
SQ Sequence 1027 AA:

Query Match 100.0%; Score 1027; DB 22; Length 1027;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1027; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFRAGPVGRGWRPAEAPRRGSPWPSPAWICWALAGCAAWAGDLPSSSRPLPP 60
DB 1 MLFRAGPVGRGWRPAEAPRRGSPWPSPAWICWALAGCAAWAGDLPSSSRPLPP 60

QY 61 COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFCASGEYLEMKNQVC 120
DB 61 COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFCASGEYLEMKNQVC 120

QY 121 SKCGEGTSLSGIKFDEWDELPAFNSIATFMDTVVGPDSRPGCCNNSWIPRGNYIE 180
DB 121 SKCGEGTSLSGIKFDEWDELPAFNSIATFMDTVVGPDSRPGCCNNSWIPRGNYIE 180

QY 181 SNRDDCTVSLIYAVHLKSGYVFEFYQVVDNNIIEFFIIONDQCEMDTTTDKWKLTND 240
DB 181 SNRDDCTVSLIYAVHLKSGYVFEFYQVVDNNIIEFFIIONDQCEMDTTTDKWKLTND 240

QY 241 GEMGSHVLMKSGTILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSCFPCKPCTFSN 300
DB 241 GEMGSHVLMKSGTILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSCFPCKPCTFSN 300

QY 301 KPGSNQVCPNTYSEKAKCIRCKDSDQSFSSSECTERPPCTTKDYFQIHTPCDEEG 360
DB 301 KPGSNQVCPNTYSEKAKCIRCKDSDQSFSSSECTERPPCTTKDYFQIHTPCDEEG 360

QY 361 KTQIMYKWIETKICREDLTDALRLPSPGKDKPCNPGFYNGSSSCHPCPPGTFSDGT 420
DB 361 KTQIMYKWIETKICREDLTDALRLPSPGKDKPCNPGFYNGSSSCHPCPPGTFSDGT 420

QY 421 KECRPCPAGTEPALGFYKWNVLPGNMKTSCFNVGNKSCDGMNGWEVAGDHIQSGAGGS 480
DB 421 KECRPCPAGTEPALGFYKWNVLPGNMKTSCFNVGNKSCDGMNGWEVAGDHIQSGAGGS 480

QY 481 DNDYLILNLHIFGFPPTSMTCATGSELGRITFFVETLCSADCVLVFWVDINRKSTNVVE 540
DB 481 DNDYLILNLHIFGFPPTSMTCATGSELGRITFFVETLCSADCVLVFWVDINRKSTNVVE 540

QY 541 SWGGTKEQAYTHIIFKNATFTTFAFORTNOGQNRREINDMWKIYSTITATNAVGVAS 600
DB 541 SWGGTKEQAYTHIIFKNATFTTFAFORTNOGQNRREINDMWKIYSTITATNAVGVAS 600

QY 601 SCRACALGSEQSGSCVCPGPGHYIEKETNOCKECPDPTYLSIHQYVGEACIPCGPGSK 660
DB 601 SCRACALGSEQSGSCVCPGPGHYIEKETNOCKECPDPTYLSIHQYVGEACIPCGPGSK 660

QY 661 NNQDHSVCSYDCFFYHEKNOILHDFSNLSSVSGIMNCPSTSKGTKYFHEFNISLGH 720
DB 661 NNQDHSVCSYDCFFYHEKNOILHDFSNLSSVSGIMNCPSTSKGTKYFHEFNISLGH 720

QY 721 EGKKNALCTNITDFTVEIVAGSDYTNLVGAFVQCSTIIPSESKGFRALSSQSIIILA 780
DB 721 EGKKNALCTNITDFTVEIVAGSDYTNLVGAFVQCSTIIPSESKGFRALSSQSIIILA 780

QY 781 DFIGVTVETTLKNIKEDMPVPVTSQIPDVHFFYKSSSTATTSCTINGRSTAVKMRCPNT 840

DB 781 DFIGVTVETTLKNIKEDMPVPVTSQIPDVHFFYKSSSTATTSCTINGRSTAVKMRCPNT 840
QY 841 KSGAGVISVPSKCPAGTCDGCTFYFLWESAEACPLCTEHDFHEIGACKRGQETLYVWN 900
DB 841 KSGAGVISVPSKCPAGTCDGCTFYFLWESAEACPLCTEHDFHEIGACKRGQETLYVWN 900
QY 901 EPKWCITKIGSLPEKKLATCETVDVFWLKVAGVGAFVAVLLVATCYFWKKNOKLEYKYSK 960
DB 901 EPKWCITKIGSLPEKKLATCETVDVFWLKVAGVGAFVAVLLVATCYFWKKNOKLEYKYSK 960
QY 961 LVMTTNSKECELPAAADSCAIMEGEDNEEBVYVSNQSLGKLSLATKEKEDIHFSVQLK 1020
DB 961 LVMTTNSKECELPAAADSCAIMEGEDNEEBVYVSNQSLGKLSLATKEKEDIHFSVQLK 1020
QY 1021 TSRSNPI 1027
DB 1021 TSRSNPI 1027

RESULT 2
AAB70255
ID AAB70255 standard; protein; 963 AA.
XX
AC AAB70255;
XX
DT 10-MAY-2001 (first entry)
XX
DE TR16-short receptor protein.
XX
KW TR16 receptor; tumour necrosis factor receptor superfamily;
KW apoptosis; inflammatory; cancer; immune; neurodegenerative.
XX
OS Unidentified.
XX
PN WO200112671-A1.
PD 22-FEB-2001.
XX
PF 10-AUG-2000; 2000WO-US21885.
PR 12-AUG-1999; 99US-0148348.
PR 13-AUG-1999; 99US-0148683.
PR 13-AUG-1999; 99US-0148870.
PR 16-AUG-1999; 99US-0148758.
PR 17-AUG-1999; 99US-0149181.
PR 18-AUG-1999; 99US-0149453.
PR 19-AUG-1999; 99US-0149498.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Young PE, Baker KP;
XX
XX
DR WPI: 2001-138754/14.
XX
PT New nucleic acid molecule encoding a TR16 tumor necrosis factor
PT receptor polypeptide, useful for the diagnosis and treatment of cancer,
PT autoimmune disorders and cardiovascular diseases -
XX
PS Claim 1; Fig 1; 286pp; English.
XX
CC The present invention relates to a TR16 receptor (tumour necrosis
CC factor receptor superfamily). The invention is useful treating
CC diseases and disorders associated with the inhibited or increased
CC apoptosis. In particular inflammatory diseases, cancers, immune and
CC neurodegenerative disorders may be treated.
XX
SQ Sequence 963 AA:

Query Match 92.8%; Score 953; DB 22; Length 963;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 953; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFRARPGVRGWRGPAEAPRGRSPWPSPAWICWALAGCAAWAGDLPSSSSRLP 60
DB 1 MLFRARPGVRGWRGPAEAPRGRSPWPSPAWICWALAGCAAWAGDLPSSSSRLP 60
QY 61 COEKDYHFEYTCDDSSGSRWRVAIPNSAVDCSGLDPPVRGKCTFSCASGEYLEMKNQVC 120
DB 61 COEKDYHFEYTCDDSSGSRWRVAIPNSAVDCSGLDPPVRGKCTFSCASGEYLEMKNQVC 120
QY 121 SKCGEGTYSLGSGIKFDEWDELPAFNSIATFMDTVVGPDSRDPDGCNNSWIPRGN 180
DB 121 SKCGEGTYSLGSGIKFDEWDELPAFNSIATFMDTVVGPDSRDPDGCNNSWIPRGN 180
QY 181 SNRDDCTVSLIYAVHLKSGYVFEFYQYVDNIIFFEFFIQQDQCEMDTTDKWKLTON 240
DB 181 SNRDDCTVSLIYAVHLKSGYVFEFYQYVDNIIFFEFFIQQDQCEMDTTDKWKLTON 240
QY 241 GEGSHSVMLKSGTNILYWRITGILMGSKAVKPVLVKNITIEGVAYTSCFCKPGTFSN 300
DB 241 GEGSHSVMLKSGTNILYWRITGILMGSKAVKPVLVKNITIEGVAYTSCFCKPGTFSN 300
QY 301 KPGSFNCQVCPRTYSEKGAKEICRCKDDSQFSGSSECTERPPCTTKDYFOIHTPCDEEG 360
DB 301 KPGSFNCQVCPRTYSEKGAKEICRCKDDSQFSGSSECTERPPCTTKDYFOIHTPCDEEG 360
QY 361 KTOIMYKWIETPKICREDLTDALRLPPSGEKKDCPCPCNPGFYNGSSCHPCPPGTFSDGT 420
DB 361 KTOIMYKWIETPKICREDLTDALRLPPSGEKKDCPCPCNPGFYNGSSCHPCPPGTFSDGT 420
QY 421 KECRCPAGTEPALGFEYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGS 480
DB 421 KECRCPAGTEPALGFEYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGS 480
QY 481 DNDYLNLNHPGFRPPTSMGTATGSELGRITFVFETLCSADCVLFYFMDINRKSTNVYE 540
DB 481 DNDYLNLNHPGFRPPTSMGTATGSELGRITFVFETLCSADCVLFYFMDINRKSTNVYE 540
QY 541 SGGTKEKQAYTHILFKNATFTTFAFORTNOGDNRRINDVMKIYTSITAINAVDGVAS 600
DB 541 SGGTKEKQAYTHILFKNATFTTFAFORTNOGDNRRINDVMKIYTSITAINAVDGVAS 600
QY 601 SCRACALGSEQSGSCVCPGPGHYIEKETNOCKECPDPTLYSLTHQVYGEACIPCGPGSK 660
DB 601 SCRACALGSEQSGSCVCPGPGHYIEKETNOCKECPDPTLYSLTHQVYGEACIPCGPGSK 660
QY 661 NNQDHSVYSDCFFYHEKENQILHYDFSNLSSVGLMNGPSTSKGTYPHFHFNISLCGH 720
DB 661 NNQDHSVYSDCFFYHEKENQILHYDFSNLSSVGLMNGPSTSKGTYPHFHFNISLCGH 720
QY 721 EGKMMALCTNNITDFTVKEIVAGSDDYTNLVGAFCQSTILIPSESKGFRAALSSQSIILA 780
DB 721 EGKMMALCTNNITDFTVKEIVAGSDDYTNLVGAFCQSTILIPSESKGFRAALSSQSIILA 780
QY 781 DTFIGVTETTLKNINIKEDMFPVPTSQIPDVHFFYKSSATTSCINGRSTAVKMCNPT 840
DB 781 DTFIGVTETTLKNINIKEDMFPVPTSQIPDVHFFYKSSATTSCINGRSTAVKMCNPT 840
QY 841 KSGAGVISVPSKPCAGTCGCTFYFLWESAECPLCTEHDHFIEGACRKGFOETLYVNN 900
DB 841 KSGAGVISVPSKPCAGTCGCTFYFLWESAECPLCTEHDHFIEGACRKGFOETLYVNN 900
QY 901 EPKWCIKGSLPEKKLATCETDVFMLKVGAGVCAFTAVLLVALTCYFVKKNQK 953
DB 901 EPKWCIKGSLPEKKLATCETDVFMLKVGAGVCAFTAVLLVALTCYFVKKNQK 953

RESULT 3
AAB48377

ID AAB48377 standard; Protein; 464 AA.

XX AAB48377;

XX AAB48377;

XX 20-APR-2001 (first entry)

XX

DE Human SEC10 protein sequence (clone ID 1795045.0.77).
XX SECX: cytostatic; gynecological; gene therapy; screening assay; human;
KW SEC10: chromosomal mapping; forensic biology; cell proliferation; cancer;
KW cell differentiation; immune associated disorder; gestational disease.
XX Homo sapiens.
XX WO200078802-A2.
XX 28-DEC-2000.
XX 23-JUN-2000; 2000WO-US17328.
XX 23-JUN-1999; 99US-0140584.
PR 20-JUL-1999; 99US-0144722.
PR 16-SEP-1999; 99US-0154520.
XX 22-JUN-2000; 2000US-0604286.
XX (CURA-) CURAGEN CORP.
XX Shinkets RA, Fernandes E, Vernet C, Yang M, Boldog FL;
PI Herrmann JL;
XX WPI; 2001-071385/08.
DR N-PSDB; AAC84891.
XX Polynucleotides encoding SECX proteins useful for treating disease
PT characterized by an aberrant level of cell proliferation and/or
PT differentiation like cancer or immune associated disorders -
XX Claim 1; Fig 10; 132pp; English.
XX The invention relates to human SECX polypeptides and polynucleotides
CC encoding them. The SECX polypeptides can be expressed by standard
CC recombinant methodology. The SECX polypeptides are useful for treating
CC or preventing a SECX-associated disorder. The invention is useful in
CC screening assays; detection assays (e.g. chromosomal mapping, cell and
CC tissue typing, forensic biology); predictive medicine (diagnostic assays,
CC prognostic assays, monitoring clinical trials, and pharmacogenomics); and
CC methods of treatment (e.g. therapeutic and prophylactic), especially
CC disorders characterized by aberrant cell proliferation and/or
CC differentiation like cancer or immune associated disorders or gestational
CC disease. The present sequence represents a SEC10 protein.
XX Sequence 464 AA:
Query Match 26.4%; Score 271; DB 22; Length 464;
Best Local Similarity 99.6%; Pred. No. 4.6e-266;
Matches 451; Conservative 0; Mismatches 0; Indels 2; Gaps 1;
QY 115 MKNQVSKCGEGTYSLGSGIKFDEWDELPAFNSIATFMDTVVGPDSRDPDGCNNSWIP 174
DB 1 MKNQVSKCGEGTYSLGSGIKFDEWDELPAFNSIATFMDTVVGPDSRDPDGCNNSWIP 60
QY 175 RGNVIESNRDDCTVSLIYAVHLKSGYVFEFYQYVDNIIFFEFFIQQDQCEMDTTDKW 234
DB 61 RGNVIESNRDDCTVSLIYAVHLKSGYVFEFYQYVDNIIFFEFFIQQDQCEMDTTDKW 120
QY 235 VKLTDNGEWSHVMKSGTNILYWRITGILMGSKAVKPVLVKNITIEGVAYTSCFPPCK 294
DB 121 VKLTDNGEWSHVMKSGTNILYWRITGILMGSKAVKPVLVKNITIEGVAYTSCFPPCK 180
QY 295 PGTFSNKPGSFNCQVCPRTYSEKGAKEICRCKDDSQFSGSSECTERPPCTTKDYFOI 352
DB 181 PGTFSNKPGSFNCQVCPRTYSEKGAKEICRCKDDSQFSGSSECTERPPCTTKDYFOI 240
QY 353 HTPCDEEGKTQIMYKWIETPKICREDLTDALRLPPSGEKKDCPCPCNPGFYNGSSCHPCP 412
DB 241 HTPCDEEGKTQIMYKWIETPKICREDLTDALRLPPSGEKKDCPCPCNPGFYNGSSCHPCP 300
QY 413 PGTFSDGTKECRPCAGTEPALGFEYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDH 472
DB 113 PGTFSDGTKECRPCAGTEPALGFEYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDH 111

```
Db 301 PGTFSDGTEKCRPCPAGTEPALGFEYKWNVLPGNNKMTSCFNVGNSKCDGMNGWEVAGDH 360
QY 473 IQSGAGSDNDYLILNLHIPGKPPPTSMTGATGSELGRITFTVFETLCSADCVLVFMVDIN 532
Db 361 IQSGAGSDNDYLILNLHIPGKPPPTSMTGATGSELGRITFTVFETLCSADCVLVFMVDIN 420
QY 533 RKSTNVVSWGGTKEKOAYTHIIFKNATFTFTW 565
Db 421 RKSTNVVSWGGTKEKOAYTHIIFKNATFTFTW 453

RESULT 4
AAB48372
ID AAB48372 standard; Protein; 411 AA.
XX
AC AAB48372;
XX
DT 20-APR-2001 (first entry)
XX
DE Human SEC5 protein sequence (clone ID 1795045.0.61).
XX
KW SEC5; cytostatic; gynecological; gene therapy; screening assay; human;
KW SEC5; chromosomal mapping; forensic biology; cell proliferation; cancer;
KW cell differentiation; immune associated disorder; gestational disease.
XX
OS Homo sapiens.
XX
PN WO200078802-A2.
XX
PD 28-DEC-2000.
XX
PF 23-JUN-2000; 2000WO-US17328.
XX
PR 23-JUN-1999; 99US-0140584.
PR 20-JUL-1999; 99US-0144722.
PR 16-SEP-1999; 99US-0154520.
PR 22-JUN-2000; 2000US-0604286.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shinkets RA, Fernandes E, Vernet C, Yang M, Boldog FL;
PI Herrmann JL;
XX
DR WPI; 2001-071385/08.
XX
N-PSDB; AAC84886.
XX
Polynucleotides encoding SECX proteins useful for treating disease
PT characterized by an aberrant level of cell proliferation and/or
PT differentiation like cancer or immune associated disorders -
XX
PS Claim 1; Fig 6; 132pp; English.
XX
The invention relates to human SECX polypeptides and polynucleotides
CC encoding them. The SECX polypeptides can be expressed by standard
CC recombinant methodology. The SECX polypeptides are useful for treating
CC or preventing a SECX-associated disorder. The invention is useful in
CC screening assays; detection assays (e.g. chromosomal mapping, cell and
CC tissue typing, forensic biology); predictive medicine (diagnostic assays,
CC prognostic assays, monitoring clinical trials, and pharmacogenomics); and
CC methods of treatment (e.g. therapeutic and prophylactic), especially
CC disorders characterized by aberrant cell proliferation and/or
CC differentiation like cancer or immune associated disorders or gestational
CC disease. The present sequence represents a SEC5 protein.
XX
SQ Sequence 411 AA;

Query Match 22.6%; Score 232; DB 22; Length 411;
Best Local Similarity 100.0%; Pred. No. 1.7e-226;
Matches 232; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 GSSECTERPPCTTKDYFQHTPCDEEGKTOIMYKWIETPKICREDLTDALRPPSGEKKDC 393
Db 169 GSSECTERPPCTTKDYFQHTPCDEEGKTOIMYKWIETPKICREDLTDALRPPSGEKKDC 228
```

```
QY 394 PCNPFGYNNNGSSSCHPCPPGTFSDGTEKCRPCPAGTEPALGFEYKWNVLPGNNKMTSCF 453
Db 229 PCNPFGYNNNGSSSCHPCPPGTFSDGTEKCRPCPAGTEPALGFEYKWNVLPGNNKMTSCF 288
QY 454 NVGNSKCDGMNGWEVAGDHIQSGAGSDNDYLILNLHIPGKPPPTSMTGATGSELGRITF 513
Db 289 NVGNSKCDGMNGWEVAGDHIQSGAGSDNDYLILNLHIPGKPPPTSMTGATGSELGRITF 348
QY 514 VFETLCSADCVLVFMVDINRKSTNVVSWGGTKEKOAYTHIIFKNATFTFTW 565
Db 349 VFETLCSADCVLVFMVDINRKSTNVVSWGGTKEKOAYTHIIFKNATFTFTW 400
```

```
RESULT 5
ABB39918
ID ABB39918 standard; Peptide; 81 AA.
XX
AC ABB39918;
XX
DT 04-FEB-2002 (first entry)
XX
DE Peptide #7424 encoded by human foetal liver single exon probe.
KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
(MOLE-) MOLECULAR DYNAMICS INC.
XX
Penn SG, Hanzel DK, Chen W, Rank DR;
XX
WPI; 2001-483447/52.
XX
Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
Claim 27; SEQ ID NO 32553; 639pp + sequence listing; English.
XX
The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a peptide encoded by a single exon
CC nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 81 AA;
```

```
Query Match 7.9%; Score 81; DB 22; Length 81;
Best Local Similarity 100.0%; Pred. No. 9.8e-74;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 578 RFINDMVKIYSITATNAVGVASSCRACALGSEQSGSCVPCPPGHIETKNOCKECP 637
Db 1 RFINDMVKIYSITATNAVGVASSCRACALGSEQSGSCVPCPPGHIETKNOCKECP 60
```

```
QY 638 DTYLSIHQVYGKACIPCGPG 658
Db 61 DTYLSIHQVYGKACIPCGPG 81

RESULT 6
ABB24471
ID ABB24471 standard; Protein; 81 AA.
AC
XX
XX
DT 23-JAN-2002 (first entry)
XX
DE Protein #6470 encoded by probe for measuring heart cell gene expression.
KW Human; gene expression; heart; microarray; vascular system;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease.
XX
XX Homo sapiens.
OS
XX WO200157274-A2.
PN
XX 09-AUG-2001.
PD
XX 30-JAN-2001; 2001WO-US00666.
PF
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR
XX 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-488899/53.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts.
PT
XX
XX Claim 15; SEQ ID No 26241; 530pp; English.
PS
XX The present invention relates to single exon nucleic acid probes for
XX measuring human gene expression in a sample derived from human heart (see
XX ABA21535-ABA41305). The present sequence is a protein encoded by one such
XX probe. The probes may be used for predicting, measuring and displaying
XX gene expression in samples derived from the human heart via microarrays.
XX By measuring gene expression, the probes are useful for predicting,
XX diagnosing, grading, staging, monitoring and prognosing diseases of the
XX human heart and vascular system e.g. cardiovascular disease,
XX hypertension, cardiac arrhythmias and congenital heart disease.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 81 AA;

Query Match 7.9%; Score 81; DB 22; Length 81;
Best Local Similarity 100.0%; Pred. No. 9.8e-74;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 578 RFINDMKIYISITATNAVGVASSCRACALGSEQSGSCVPCPPGGHYIEKETNOCKECPP 637
Db 1 RFINDMKIYISITATNAVGVASSCRACALGSEQSGSCVPCPPGGHYIEKETNOCKECPP 60

QY 638 DTYLSIHQVYGKACIPCGPG 658
Db 61 DTYLSIHQVYGKACIPCGPG 81

RESULT 7
AAM60663
ID AAM60663 standard; Protein; 81 AA.
AC
XX
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 32768.
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer.
XX
XX Homo sapiens.
OS
XX WO200157275-A2.
PN
XX 09-AUG-2001.
PD
XX 30-JAN-2001; 2001WO-US00667.
PF
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR
XX 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains.
PT
XX
XX Example 4; SEQ ID NO: 32768; 650pp + Sequence Listing; English.
PS
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is a protein encoded by one of
XX the probes of the invention.
XX
XX Sequence 81 AA;

Query Match 7.9%; Score 81; DB 22; Length 81;
Best Local Similarity 100.0%; Pred. No. 9.8e-74;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 578 RFINDMKIYISITATNAVGVASSCRACALGSEQSGSCVPCPPGGHYIEKETNOCKECPP 637
Db 1 RFINDMKIYISITATNAVGVASSCRACALGSEQSGSCVPCPPGGHYIEKETNOCKECPP 60

QY 638 DTYLSIHQVYGKACIPCGPG 658
Db 61 DTYLSIHQVYGKACIPCGPG 81

RESULT 8
AAM73335
ID AAM73335 standard; Protein; 81 AA.
AC
XX
XX
DT 06-NOV-2001 (first entry)
XX
XX
```

```
DE Human bone marrow expressed probe encoded protein SEQ ID NO: 33641.
XX Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX Homo sapiens.
XX WO200157276-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US00668.
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow.
XX Example 4; SEQ ID NO: 33641; 658pp + Sequence Listing; English.
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention.
XX Sequence 81 AA;
XX
XX Query Match 7.9%; Score 81; DB 22; Length 81;
XX Best Local Similarity 100.0%; Pred. No. 9.8e-74;
XX Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 578 REINDMKIYISITATNAVGVASSCRACALGSEQSGSCVPCPPGGHYIEKETNCKECP 637
XX 1 REINDMKIYISITATNAVGVASSCRACALGSEQSGSCVPCPPGGHYIEKETNCKECP 60
XX
XX QY 638 DTYLSIHQVYGKACIPCGPG 658
XX 61 DTYLSIHQVYGKACIPCGPG 81
XX
XX RESULT 9
XX AAM33535
XX ID AAM33535 standard; Protein; 81 AA.
XX AC AAM33535;
XX XX
XX DT 17-OCT-2001 (first entry)
XX DE
XX PEptide #7572 encoded by probe for measuring placental gene expression.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX genetic disorder.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX XX
XX PD 09-AUG-2001.
XX XX
```

```
PF 30-JAN-2001; 2001WO-US00663.
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-48897/53.
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta.
XX Claim 27; SEQ ID NO 33804; 654pp; English.
XX The present invention relates to single exon nucleic acid probes (SENP:
XX see AAI31315-AAI57546). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for producing a microarray for
XX predicting, measuring and displaying gene expression in samples derived
XX from human placenta. The probes are useful for antenatal diagnosis of
XX human genetic disorders.
XX Sequence 81 AA;
XX
XX Query Match 7.9%; Score 81; DB 22; Length 81;
XX Best Local Similarity 100.0%; Pred. No. 9.8e-74;
XX Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 578 REINDMKIYISITATNAVGVASSCRACALGSEQSGSCVPCPPGGHYIEKETNCKECP 637
XX 1 REINDMKIYISITATNAVGVASSCRACALGSEQSGSCVPCPPGGHYIEKETNCKECP 60
XX
XX QY 638 DTYLSIHQVYGKACIPCGPG 658
XX 61 DTYLSIHQVYGKACIPCGPG 81
XX
XX RESULT 10
XX ABG43186
XX ID ABG43186 standard; Peptide; 81 AA.
XX AC ABG43186;
XX XX
XX DT 19-AUG-2002 (first entry)
XX XX
XX DE Human peptide encoded by genome-derived single exon probe SEQ ID 32851.
XX KW Human; single exon probe; asthma; lung cancer; COPD; ILD;
XX chronic obstructive pulmonary disease; interstitial lung disease;
XX familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemorrhage;
XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
XX pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
XX primary ciliary dyskinesia; pulmonary hypertension;
XX hyaline membrane disease.
XX OS Homo sapiens.
XX XX
XX PN WO200186003-A2.
XX XX
XX PD 15-NOV-2001.
XX XX
XX 30-JAN-2001; 2001WO-US00665.
XX PF
XX 04-FEB-2000; 2000US-180312P.
XX PR 26-MAY-2000; 2000US-207456P.
XX PR
```

PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-234687P.
PR 27-SEP-2000; 2000US-236359P.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2002-114183/15.
XX
PT Spatially-addressable set of single exon nucleic acid probes, used to
PT measure gene expression in human lung samples -
XX
PS Claim 27; SEQ ID No 32851; 634pp; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements or the 12387 open reading frames derived from the 12614
CC probes. Also included are a microarray comprising the novel set of
CC probes; the novel set of probes which hybridize at high stringency to a
CC nucleic acid expressed in the human lung; measuring gene expression to a
CC sample derived from human lung, comprising (a) contacting the array with
CC a collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of
CC the array; identifying exons in a eukaryotic genome, comprising
CC (a) algorithmically predicting at least one exon from genomic sequences
CC of the eukaryote; and (b) detecting specific hybridisation of detectably
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC having a fragment identical to the predicted exon, the probe is included
CC in the above mentioned microarray; assigning exons to a single gene,
CC comprising (a) identifying exons from genomic sequence by the method
CC above and (b) measuring the expression of each of the exons in several
CC tissues and/or cell types using hybridisation to a single exon
CC microarrays having a probe with the exon, where a common pattern of
CC expression of the exons in the tissues and/or cell types indicates that
CC the exons should be assigned to a single gene; a peptide comprising one
CC of 12011 sequences, mentioned in the specification, or encoded by the
CC probes/open reading frames (ORF). The probes are used for gene
CC expression analysis, and for identifying exons in a gene, particularly
CC using human lung derived mRNA and for the study of lung diseases
CC such as asthma, lung cancer, chronic obstructive pulmonary disease
CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
CC and hyaline membrane disease. The present sequence is a peptide/protein
CC encoded by a single exon probe of the invention.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 81 AA;

Query Match 7.9%; Score 81; DB 23; Length 81;
Best Local Similarity 100.0%; Pred. No. 9.8e-74;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 578 RFINDMWKIYITATNAVGVASSRACALGSEQSGSCVPCPPGHYIPKTNQCKECP 637
DB 1 RFINDMWKIYITATNAVGVASSRACALGSEQSGSCVPCPPGHYIPKTNQCKECP 60

OY 638 DYLSLHVQYVKGKACIPCGP 658
DB 61 DYLSLHVQYVKGKACIPCGP 81

RESULT 11
AAB70285
ID AAB70285 standard; peptide; 78 AA.
XX
AC AAB70285;
XX
DT 10-MAY-2001 (first entry)
XX
DE Peptide #29.
XX
KW TR16 receptor; tumour necrosis factor receptor superfamily;
KW apoptosis; inflammatory; cancer; immune; neurodegenerative.
XX
OS Unidentified.
XX
PN WO200112671-A1.
XX
PD 22-FEB-2001.
XX
PF 10-AUG-2000; 2000WO-US21885.
XX
PR 12-AUG-1999; 99US-0148348.
PR 13-AUG-1999; 99US-0148683.
PR 13-AUG-1999; 99US-0148870.
PR 16-AUG-1999; 99US-0148758.
PR 17-AUG-1999; 99US-0149181.
PR 18-AUG-1999; 99US-0149453.
PR 19-AUG-1999; 99US-0149498.
XX
(HUMA-) HUMAN GENOME SCI INC.
PA
PI Ruben SM, Young PE, Baker KP;
XX
DR WPI; 2001-138754/14.
XX
PT New nucleic acid molecule encoding a TR16 tumor necrosis factor
PT receptor polypeptide, useful for the diagnosis and treatment of cancer,
PT autoimmune disorders and cardiovascular diseases -
XX
PS Disclosure; Page 81; 286pp; English.
XX
CC The present invention relates to a TR16 receptor (tumour necrosis
CC factor receptor superfamily). The invention is useful treating
CC diseases and disorders associated with the inhibited or increased
CC apoptosis. In particular inflammatory diseases, cancers, immune and
CC neurodegenerative disorders may be treated.
XX
SQ Sequence 78 AA;
Query Match 7.6%; Score 78; DB 22; Length 78;
Best Local Similarity 100.0%; Pred. No. 1.1e-70;
Matches 78; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 950 KNOKLEYKYSKLVMTTNSKECELPAAADSCAIMEGDEEVEVYVSNKQSLGKLKSLATKE 1009
DB 1 KNOKLEYKYSKLVMTTNSKECELPAAADSCAIMEGDEEVEVYVSNKQSLGKLKSLATKE 60
OY 1010 KEDHFESVOLKTSRSPNI 1027
DB 61 KEDHFESVOLKTSRSPNI 78

RESULT 12
AAB70281
ID AAB70281 standard; peptide; 74 AA.
XX
AC AAB70281;
XX
DT 10-MAY-2001 (first entry)
XX
DE Peptide #25.
XX
KW TR16 receptor; tumour necrosis factor receptor superfamily;

KW apoptosis; inflammatory; cancer; immune; neurodegenerative.
XX Unidentified.
OS WO200112671-A1.
XX 22-FEB-2001.
XX 10-AUG-2000; 2000WO-US21885.
XX 12-AUG-1999; 99US-0148348.
XX 13-AUG-1999; 99US-0148683.
XX 13-AUG-1999; 99US-0148870.
XX 16-AUG-1999; 99US-0148758.
XX 17-AUG-1999; 99US-0149181.
XX 18-AUG-1999; 99US-0149453.
XX 19-AUG-1999; 99US-0149498.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, Young PE, Baker KP;
XX WPI; 2001-138754/14.
XX New nucleic acid molecule encoding a TRL6 tumor necrosis factor
XX receptor polypeptide, useful for the diagnosis and treatment of cancer,
XX autoimmune disorders and cardiovascular diseases -
XX Disclosure; Page 81; 286pp; English.
XX The present invention relates to a TRL6 receptor (tumour necrosis
XX factor receptor superfamily). The invention is useful treating
XX diseases and disorders associated with the inhibited or increased
XX apoptosis. In particular inflammatory diseases, cancers, immune and
XX neurodegenerative disorders may be treated.
XX Sequence 74 AA;
XX
XX Query Match 7.2%; Score 74; DB 22; Length 74;
XX Best Local Similarity 100.0%; Pred. No. 1.2e-66;
XX Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 338 CTERPCTTKDYFQIHTPCDEEGTKQIMYKTEPKICREDLTDALRLPSPGKDKCPCPN 397
XX DB 1 CTERPCTTKDYFQIHTPCDEEGTKQIMYKTEPKICREDLTDALRLPSPGKDKCPCPN 60
XX
XX QY 398 PCFYNNNGSSSCHPC 411
XX DB 61 PCFYNNNGSSSCHPC 74
XX
XX RESULT 13
XX AAM72925
XX ID AAM72925 standard; Protein; 64 AA.
XX AC AAM72925;
XX XX
XX 06-NOV-2001 (first entry)
XX Human bone marrow expressed probe encoded protein SEQ ID NO: 33231.
XX Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma.
XX Homo sapiens.
XX WO200157276-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US00668.
XX 04-FEB-2000; 2000US-0180312.
XX
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human bone marrow -
XX Example 4; SEQ ID NO: 33231; 658pp + Sequence Listing; English.
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention.
XX Sequence 64 AA;
XX
XX Query Match 6.2%; Score 64; DB 22; Length 64;
XX Best Local Similarity 100.0%; Pred. No. 1.5e-56;
XX Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 105 FSCASGEYLEMKNQVCSKCGEGTYSIGSGIKFDEWDELPAFSGSNATFMDTVVGFSDSRP 164
XX DB 1 FSCASGEYLEMKNQVCSKCGEGTYSIGSGIKFDEWDELPAFSGSNATFMDTVVGFSDSRP 60
XX
XX QY 165 DGCN 168
XX DB 61 DGCN 64
XX
XX RESULT 14
XX ABG42760
XX ID ABG42760 standard; Peptide; 64 AA.
XX AC ABG42760;
XX XX
XX 19-AUG-2002 (first entry)
XX Human peptide encoded by genome-derived single exon probe SEQ ID 32425.
XX
XX Human; single exon probe; asthma; lung cancer; COPD; ILD;
XX chronic obstructive pulmonary disease; interstitial lung disease;
XX familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemostiderosis;
XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karagenier syndrome;
XX pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
XX primary ciliary dyskinesia; pulmonary hypertension;
XX hyaline membrane disease.
XX Homo sapiens.
XX WO200186003-A2.
XX 15-NOV-2001.
XX 30-JAN-2001; 2001WO-US00665.
XX 04-FEB-2000; 2000US-180312P.
XX 26-MAY-2000; 2000US-207456P.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-234687P.

```
PR 27-SEP-2000; 2000US-236359P.
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2002-114183/15.
XX Spatially-addressable set of single exon nucleic acid probes, used to
XX measure gene expression in human lung samples -
XX Claim 27; SEQ ID NO 32425; 634pp; English.
XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human lung comprising single exon nucleic acid probes having one of
XX 12614 nucleic acid sequences mentioned in the specification, or their
XX complements or the 12387 open reading frames derived from the 12614
XX probes. Also included are a microarray comprising the novel set of
XX probes; the novel set of probes which hybridise at high stringency to a
XX nucleic acid expressed in the human lung; measuring gene expression in a
XX sample derived from human lung, comprising (a) contacting the array with
XX a collection of detectably labeled nucleic acids derived from human lung
XX mRNA, and (b) measuring the label detectably bound to each probe of
XX the array; identifying exons in a eukaryotic genome, comprising
XX (a) algorithmically predicting at least one exon from genomic sequences
XX of the eukaryote; and (b) detecting specific hybridisation of detectably
XX labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
XX having a fragment identical to the predicted exon, the probe is included
XX in the above mentioned microarray; assigning exons to a single gene,
XX comprising (a) identifying exons from genomic sequences by the method
XX above and (b) measuring the expression of each of the exons in several
XX tissues and/or cell types using hybridisation to a single exon
XX microarrays having a probe with the exon, where a common pattern of
XX expression of the exons in the tissues and/or cell types indicates that
XX the exons should be assigned to a single gene; a peptide comprising one
XX of 12011 sequences, mentioned in the specification, or encoded by the
XX probes/open reading frames (ORF). The probes are used for gene
XX expression analysis, and for identifying exons in a gene, particularly
XX using human lung derived mRNA and for the study of lung diseases
XX such as asthma, lung cancer, chronic obstructive pulmonary disease
XX (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
XX fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
XX Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
XX haemostiderosis, pulmonary histiocytosis, lymphangioma, lymphomatosis,
XX pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
XX pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
XX and hyaline membrane disease. The present sequence is a peptide/protein
XX encoded by a single exon probe of the invention.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 64 AA;
SQ
Query Match 6.2%; Score 64; DB 23; Length 64;
Best Local Similarity 100.0%; Pred. No. 1.5e-56;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 105 FSCASGEYLEMKNQVCKGEGTSLGSGIKFDEWDELPAFSGNSIATFMDTVVGPDSRSP 164
DB 1 FSCASGEYLEMKNQVCKGEGTSLGSGIKFDEWDELPAFSGNSIATFMDTVVGPDSRSP 60
QY 165 DGCN 168
DB 61 DGCN 64
RESULT 15
ABB38686
ID ABB38686 standard; Peptide; 60 AA.
```

```
XX ABB38686;
XX 04-FEB-2002 (first entry)
XX Peptide #6192 encoded by human foetal liver single exon probe.
XX Human; foetal liver; gene expression; single exon nucleic acid probe.
XX Homo sapiens.
XX WO200157277-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000669.
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver -
XX Claim 27; SEQ ID NO 31321; 639pp + sequence listing; English.
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX foetal liver. The present sequence is a peptide encoded by a single exon
XX nucleic acid probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 60 AA;
SQ
Query Match 5.8%; Score 60; DB 22; Length 60;
Best Local Similarity 100.0%; Pred. No. 1.6e-52;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 362 TQIMYKWIPEKICREDLTDAILRPPSGEKKDCPPCNPNGFYNNSSCHPCPPGTFSDGTK 421
DB 1 TQIMYKWIPEKICREDLTDAILRPPSGEKKDCPPCNPNGFYNNSSCHPCPPGTFSDGTK 60
Search completed: May 12, 2003, 13:38:54
Job time : 38.1256 secs
```


GenCore version 5.1.5
Copyright (C) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 13:23:08 ; Search time 33.8744 Seconds
(without alignments)
3788.125 Million cell updates/sec

Title: US-10-073-333A-2
Perfect score: 963
Sequence: 1 MLFRAGPVRGWRGPAEA.....TCYFWKKKQKKKTLILFN 963

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	963	100.0	963	22	TR16-short receptor
2	953	99.0	1027	22	TR16-long receptor
3	271	28.1	464	22	Human SEC10 protein
4	232	24.1	411	22	Human SEC5 protein
5	81	8.4	81	22	Peptide #7424 enco
6	81	8.4	81	22	Protein #6470 enco
7	81	8.4	81	22	Human brain expres
8	81	8.4	81	22	Human bone marrow
9	81	8.4	81	22	Peptide #7572 enco
10	81	8.4	81	23	Human peptide enco

11	74	7.7	74	22	AA70281	Peptide #25. Unide
12	64	6.6	64	22	AA702925	Human bone marrow
13	64	6.6	64	23	ABG42760	Human peptide enco
14	60	6.2	60	22	AB838686	Peptide #6192 enco
15	60	6.2	60	22	AB823759	Protein #5758 enco
16	60	6.2	60	22	AA553318	Human brain expres
17	60	6.2	60	22	AA701867	Human bone marrow
18	60	6.2	60	22	AA32149	Peptide #6186 enco
19	60	6.2	60	23	ABG41680	Human peptide enco
20	15	1.6	372	22	AA85768	Human seven-transm
21	15	1.6	383	22	AA83853	Amino acid sequenc
22	15	1.6	870	22	AA83851	Amino acid sequenc
23	15	1.6	911	22	AA83850	Amino acid sequenc
24	15	1.6	1001	22	AA835333	Human TRI3 recepto
25	15	1.6	1013	21	AA26179	Human CASB619 prot
26	15	1.6	1013	22	AAU12190	Human PRO4985 poly
27	15	1.6	1013	22	AA83845	Amino acid sequenc
28	11	1.1	52	22	AA83848	Peptide fragment o
29	11	1.1	105	21	AA26180	Human CASB619 prot
30	11	1.1	209	22	AA83852	Amino acid sequenc
31	11	1.1	495	20	AA559972	Human endometrium
32	11	1.1	750	22	AA835328	Human TRI3 recepto
33	9	0.9	212	22	ABG04843	Novel human diagno
34	9	0.9	708	22	AA891270	C. glutamicum prote
35	9	0.9	1487	19	AAW76719	Thermococcus 9N2 m
36	9	0.9	1487	19	AAW76720	Thermococcus 9N2 m
37	8	0.8	8	22	AA70257	Peptide #1. Unide
38	8	0.8	8	22	AA70258	Peptide #2. Unide
39	8	0.8	8	22	AA70259	Peptide #3. Unide
40	8	0.8	8	22	AA70260	Peptide #4. Unide
41	8	0.8	8	22	AA70261	Peptide #5. Unide
42	8	0.8	8	22	AA70262	Peptide #6. Unide
43	8	0.8	8	22	AA70263	Peptide #7. Unide
44	8	0.8	8	22	AA70264	Peptide #8. Unide
45	8	0.8	8	22	AA70265	Peptide #9. Unide

ALIGNMENTS

RESULT 1
AA70255
ID AAB70255 standard; protein; 963 AA.
AC AAB70255;
XX
DT 10-MAY-2001 (first entry)
DE
DE TR16-short receptor protein.
DE
DE TR16 receptor; tumour necrosis factor receptor superfamily;
KW apoptosis; inflammatory; cancer; immune; neurodegenerative.
KW
XX Unidentified.
XX
XX WO200112671-A1.
XX
PD 22-FEB-2001.
XX
XX 10-AUG-2000; 2000WO-US21885.
XX
XX 12-AUG-1999; 99US-0148348.
PR 13-AUG-1999; 99US-0148683.
PR 13-AUG-1999; 99US-0148870.
PR 16-AUG-1999; 99US-0148758.
PR 17-AUG-1999; 99US-0149181.
PR 18-AUG-1999; 99US-0149453.
PR 19-AUG-1999; 99US-0149498.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX Ruben SM, Young PE, Baker KP;
XX

DR WPI: 2001-138754/14.
XX New nucleic acid molecule encoding a TR16 tumor necrosis factor
PT receptor polypeptide, useful for the diagnosis and treatment of cancer,
PT autoimmune disorders and cardiovascular diseases -
XX
PS Claim 1; Fig 1; 286pp; English.
XX The present invention relates to a TR16 receptor (tumour necrosis
CC factor receptor superfamily). The invention is useful treating
CC diseases and disorders associated with the inhibited or increased
CC apoptosis. In particular inflammatory diseases, cancers, immune and
CC neurodegenerative disorders may be treated.
XX
XX Sequence 963 AA:
SQ Query Match 100.0%; Score 963; DB 22; Length 963;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 963; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFRARGPVGRGWRPAEAPRRGRSPWSPAWICWALAGCOAAWAGDLPSSSSRPLPP 60
DB 1 MLFRARGPVGRGWRPAEAPRRGRSPWSPAWICWALAGCOAAWAGDLPSSSSRPLPP 60

QY 61 COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120
DB 61 COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120

QY 121 SKCGEGTSLGSGIKFDEWDELPAFNSIATFMDTVVGPDSRDPGDCNNSWIPRGNIIE 180
DB 121 SKCGEGTSLGSGIKFDEWDELPAFNSIATFMDTVVGPDSRDPGDCNNSWIPRGNIIE 180

QY 181 SNRDDCTVSLIYAVHLKSGYFFEFYQYVDNNIFFEFTQNDQCQEMDITTDKWKLTDN 240
DB 181 SNRDDCTVSLIYAVHLKSGYFFEFYQYVDNNIFFEFTQNDQCQEMDITTDKWKLTDN 240

QY 241 GEMGSHVLMKSGTNILYWRITGILMGSAKVPVLVKNITIEGVAYTSECFCKPGTFSN 300
DB 241 GEMGSHVLMKSGTNILYWRITGILMGSAKVPVLVKNITIEGVAYTSECFCKPGTFSN 300

QY 301 KPGSFNCQVCPNTYSEKAKCIRCKDDQSFSGSSECTERPCTTKDYFQIHTPCDEEG 360
DB 301 KPGSFNCQVCPNTYSEKAKCIRCKDDQSFSGSSECTERPCTTKDYFQIHTPCDEEG 360

QY 361 KTOIMYKWIETPKICREDLTDALRLPSPGSKKDCPCPNPGFYNNSSCHPCPGTFSOGT 420
DB 361 KTOIMYKWIETPKICREDLTDALRLPSPGSKKDCPCPNPGFYNNSSCHPCPGTFSOGT 420

QY 421 KECRCPAGTEPALGFEYKWNVLPONMKTSCFNVGNSKCDGMNGWEVAGDHIQSCAGGS 480
DB 421 KECRCPAGTEPALGFEYKWNVLPONMKTSCFNVGNSKCDGMNGWEVAGDHIQSCAGGS 480

QY 481 DNDYLILNLHIPGFKPPTSMGTATGSELGRITFVFTLCSADCLVYFMDINRKSTNVVE 540
DB 481 DNDYLILNLHIPGFKPPTSMGTATGSELGRITFVFTLCSADCLVYFMDINRKSTNVVE 540

QY 541 SWGTEKQAYTHIFKNAFTFTWAFQRTNOCQDNRFINDMVKIYSTATNAVGVAS 600
DB 541 SWGTEKQAYTHIFKNAFTFTWAFQRTNOCQDNRFINDMVKIYSTATNAVGVAS 600

QY 601 SCRACALGSQSSSCVPCPGPHYIEKETNOCKECPDVTLSHQVYGEACIPCGPGSK 660
DB 601 SCRACALGSQSSSCVPCPGPHYIEKETNOCKECPDVTLSHQVYGEACIPCGPGSK 660

QY 661 NNODHSVCYSDCFYHEKNOILHYDFSNLSSVGLMNGPSFTSKGTKYHFFNISLCGH 720
DB 661 NNODHSVCYSDCFYHEKNOILHYDFSNLSSVGLMNGPSFTSKGTKYHFFNISLCGH 720

QY 721 EGKMACTNNITDFTVKEIVAGSDDDYTNLVGAFVCOSTIIPSESKGFRAALSSQSIILA 780
DB 721 EGKMACTNNITDFTVKEIVAGSDDDYTNLVGAFVCOSTIIPSESKGFRAALSSQSIILA 780

QY 781 DTFIGVTVEITLKNINIKEDMFVPVTSQIPDVHFFYKSSATTSCINGRSTAVKMRNCPT 840

DB 781 DTFIGVTVEITLKNINIKEDMFVPVTSQIPDVHFFYKSSATTSCINGRSTAVKMRNCPT 840
QY 841 KSGAGVISVPSKCPAGTCDGCTFYFLWESAEACPLCTEHDHFHEIEGACKRGFQETLYVN 900
DB 841 KSGAGVISVPSKCPAGTCDGCTFYFLWESAEACPLCTEHDHFHEIEGACKRGFQETLYVN 900
QY 901 EPKWCIRKIGISLPEKKLATCETVDFWLKVGAGVGAFTAVLLVALTCYFWKKNOKKKKTILN 960
DB 901 EPKWCIRKIGISLPEKKLATCETVDFWLKVGAGVGAFTAVLLVALTCYFWKKNOKKKKTILN 960
QY 961 LFN 963
DB 961 LFN 963

RESULT 2
AAB70256
ID AAB70256 standard; protein: 1027 AA.
XX
XX AAB70256;
XX AC
XX DT 10-MAY-2001 (first entry)
XX DE TR16-long receptor protein.
XX KW TR16 receptor; tumour necrosis factor receptor superfamily;
XX KW apoptosis; inflammatory; cancer; immune; neurodegenerative.
XX OS Unidentified.
XX PN WO200112671-A1.
XX PD 22-FEB-2001.
XX PF 10-AUG-2000; 2000WO-US21885.
XX PR 12-AUG-1999; 99US-0148348.
XX PR 13-AUG-1999; 99US-0148683.
XX PR 13-AUG-1999; 99US-0148870.
XX PR 16-AUG-1999; 99US-0148758.
XX PR 17-AUG-1999; 99US-0149181.
XX PR 18-AUG-1999; 99US-0149453.
XX PR 19-AUG-1999; 99US-0149498.
XX (HUMA-) HUMAN GENOME SCI INC.
XX PA Ruben SM, Young PE, Baker KP;
XX PI WPI: 2001-138754/14.
XX DR New nucleic acid molecule encoding a TR16 tumor necrosis factor
XX receptor polypeptide, useful for the diagnosis and treatment of cancer,
XX autoimmune disorders and cardiovascular diseases -
XX PS Disclosure; Fig 4; 286pp; English.
XX The present invention relates to a TR16 receptor (tumour necrosis
XX factor receptor superfamily). The invention is useful treating
XX diseases and disorders associated with the inhibited or increased
XX apoptosis. In particular inflammatory diseases, cancers, immune and
XX neurodegenerative disorders may be treated.
XX
XX Sequence 1027 AA;
SQ Query Match 99.0%; Score 953; DB 22; Length 1027;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 953; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFRARGPVGRGWRPAEAPRRGRSPWSPAWICWALAGCOAAWAGDLPSSSSRPLPP 60
DB 1 MLFRARGPVGRGWRPAEAPRRGRSPWSPAWICWALAGCOAAWAGDLPSSSSRPLPP 60

Db 361 IQSGAGSDNDYLILNLHIPGFKPPTSMGTGATGSELGRITFVFETLCSADCVLYFMWDIN 420

Qy 533 RKSTNVVSGWGTREKQAYTHIIFKNATFTFTW 565
|||||

Db 421 RKSTNVVSGWGTREKQAYTHIIFKNATFTFTW 453
|||||

RESULT 4
AAB48372
ID AAB48372 standard; Protein; 411 AA.
XX AAB48372;
AC
XX
XX
DT 20-APR-2001 (first entry)
XX
XX Human SEC5 protein sequence (clone ID 1795045.0.61).
XX
XX SEC5; cytosolic; gynecological; gene therapy; screening assay; human;
KW SEC5; chromosomal mapping; forensic biology; cell proliferation; cancer;
KW cell differentiation; immune associated disorder; gestational disease.
XX
XX Homo sapiens.
XX
XX WO200078802-A2.
XX
XX 28-DEC-2000.
XX
XX 23-JUN-2000; 2000WO-US17328.
XX
XX 23-JUN-1999; 99US-0140584.
PR 20-JUL-1999; 99US-0144722.
PR 16-SEP-1999; 99US-0154520.
PR 22-JUN-2000; 2000US-0604286.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shinkets RA, Fernandes E, Vernet C, Yang M, Boldog FL;
PI Herrmann JL;
XX
XX WPI: 2001-071385/08.
DR N-PSDB; AAC84886.
XX
XX Polynucleotides encoding SECX proteins useful for treating disease
PT characterized by an aberrant level of cell proliferation and/or
PT differentiation like cancer or immune associated disorders -
XX
XX Claim 1; Fig 6; 132pp; English.
XX
XX The invention relates to human SECX polypeptides and polynucleotides
CC encoding them. The SECX polypeptides can be expressed by standard
CC recombinant methodology. The SECX polypeptides are useful for treating
CC or preventing a SECX-associated disorder. The invention is useful in
CC screening assays; detection assays (e.g. chromosomal mapping, cell and
CC tissue typing, forensic biology); predictive medicine (diagnostic assays,
CC prognostic assays, monitoring clinical trials, and pharmacogenomics); and
CC methods of treatment (e.g. therapeutic and prophylactic), especially
CC disorders characterized by aberrant cell proliferation and/or
CC differentiation like cancer or immune associated disorders or gestational
CC disease. The present sequence represents a SEC5 protein.
XX
XX Sequence 411 AA;

Query Match 24.1%; Score 232; DB 22; Length 411;
Best Local Similarity 100.0%; Pred. No. 3.7e-225;
Matches 232; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 334 GSSECTERPCPTKDYFOIHTPCDEBGTQIMYKWIPIKICREDTDAIRLPPSGEKKDC 393
|||||

Db 169 GSSECTERPCPTKDYFOIHTPCDEBGTQIMYKWIPIKICREDTDAIRLPPSGEKKDC 228
|||||

Qy 394 PCNPGFYNNSSSCHPCPPGTFSDGTEKRCPCPAGTEPAGLGFYKWNVLPGNMKTSF 453
|||||

Db 229 PCNPGFYNNSSSCHPCPPGTFSDGTEKRCPCPAGTEPAGLGFYKWNVLPGNMKTSF 288
|||||

Qy 454 NVGNSKCDMGNGWEVAGDHIOGAGSGSDNDYLILNLHIPGFKPPTSMGTGATGSELGRITF 513
|||||

Db 289 NVGNSKCDMGNGWEVAGDHIOGAGSGSDNDYLILNLHIPGFKPPTSMGTGATGSELGRITF 348
|||||

Qy 514 VFETLCSADCVLYFMWDINRKSTNVVSGWGTREKQAYTHIIFKNATFTFTW 565
|||||

Db 349 VFETLCSADCVLYFMWDINRKSTNVVSGWGTREKQAYTHIIFKNATFTFTW 400
|||||

RESULT 5
ABB39918
ID ABB39918 standard; Peptide; 81 AA.
XX ABB39918;
AC
XX
XX
DT 04-FEB-2002 (first entry)
XX
XX Peptide #7424 encoded by human foetal liver single exon probe.
DE
XX Human; foetal liver; gene expression; single exon nucleic acid probe.
KW
XX Homo sapiens.
OS
XX WO200157277-A2.
PN
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00669.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.

PA Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI: 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human foetal liver -
XX
XX Claim 27; SEQ ID NO 32553; 639pp + sequence listing; English.
PS
XX The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC foetal liver. The present sequence is a peptide encoded by a single exon
CC nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Qy 578 RFINDMVKIYSITATNAVGVASSCRACALGSESGSSCVPCPPGHYIERETNCKECP 637
|||||

Db 1 RFINDMVKIYSITATNAVGVASSCRACALGSESGSSCVPCPPGHYIERETNCKECP 60
|||||

Qy 638 DTLSLIIHQVYKKEACIPCGPG 658
|||||

Db 61 DTLSLIIHQVYKKEACIPCGPG 81
|||||

```

XX AC AAM60663;
XX DT 05-NOV-2001 (first entry)
XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 32768.
XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 32768.
XX KW Human; brain expressed exon; gene expression analysis; probe;
XX KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
XX KW epilepsy; cancer.
XX OS Homo sapiens.
XX PN WO200157275-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00667.
XX PP 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PX Penn SG, Hanzel DK, Chen W, Rank DR;
XX PI WPI; 2001-483446/52.
XX DR Single exon nucleic acid probes for analyzing gene expression in human
XX PT brains
XX PT Example 4; SEQ ID NO: 32768; 650pp + Sequence Listing; English.
XX PS The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC brain. They can be used to measure gene expression in brain cell samples,
XX CC which may enable the diagnosis and improved treatment of nervous system
XX CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX CC epilepsy and cancers. The present sequence is a protein encoded by one of
XX CC the probes of the invention.
XX SQ Sequence 81 AA:
XX
XX Query Match 8.4%; Score 81; DB 22; Length 81;
XX Best Local Similarity 100.0%; Pred. No. 2.7e-73;
XX Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
XX
XX QY 578 RFINDMWKIYTSITATNAVDGVASSCRACALGSEQSGSCVPCPPGGHYIEKETNQCPCPP 637
XX DB 1 RFINDMWKIYTSITATNAVDGVASSCRACALGSEQSGSCVPCPPGGHYIEKETNQCPCPP 60
XX
XX QY 638 DTYLSIHQVYGKEAICPGPG 658
XX DB 61 DTYLSIHQVYGKEAICPGPG 81
XX
XX RESULT 8
XX AAM73335
XX ID AAM73335 standard; Protein; 81 AA.
XX AC AAM73335;
XX XX
XX DT 06-NOV-2001 (first entry)
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 33641.
XX DE Human; bone marrow expressed exon; gene expression analysis; probe;
XX KW microarray; cancer; leukaemia; lymphoma; myeloma.

```

```

XX OS Homo sapiens.
XX PN WO200157276-A2.
XX XX
XX PD 09-AUG-2001.
XX XX
XX PF 30-JAN-2001; 2001WO-US00668.
XX XX
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX XX
XX DR WPI; 2001-488900/53.
XX XX
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human bone marrow.
XX XX
XX PS Example 4; SEQ ID NO: 33641; 658pp + Sequence Listing; English.
XX XX
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC bone marrow. They can be used to measure gene expression in bone marrow
XX CC samples, which may enable the improved diagnosis and treatment of cancers
XX CC such as lymphoma, leukaemia and myeloma. The present sequence is a
XX CC protein encoded by one of the probes of the invention.
XX XX
XX SQ Sequence 81 AA;

Query Match      8.4%; Score 81; DB 22; Length 81;
Best Local Similarity 100.0%; Pred. No. 2.7e-73;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 578 RFINDMWKIYSITATNAVDGVASSCRACALGSGSGSCVPCPPGHHYIEKETNCKECP 637
DB 1 RFINDMWKIYSITATNAVDGVASSCRACALGSGSGSCVPCPPGHHYIEKETNCKECP 60

QY 638 DTYLSTHGVYKREACIPCGPG 658
DB 61 DTYLSTHGVYKREACIPCGPG 81

RESULT 9
AAM33535
ID AAM33535 standard; Protein; 81 AA.
AC AAM33535;
XX XX
XX DT 17-OCT-2001 (first entry)
XX XX
XX DE Peptide #7572 encoded by probe for measuring placental gene expression.
XX XX
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX XX
XX PD 09-AUG-2001.
XX XX
XX PF 30-JAN-2001; 2001WO-US00663.
XX XX
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.

```

```

PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX XX
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX XX
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX XX
XX DR WPI; 2001-488900/53.
XX XX
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human placenta.
XX XX
XX PS Claim 27; SEQ ID NO 33804; 654pp; English.
XX XX
XX CC The present invention relates to single exon nucleic acid probes (SENP;
XX CC see AAI31315-AAI57546). The present sequence is a peptide encoded by one
XX CC such probe. The probes are useful for producing a microarray for
XX CC predicting, measuring and displaying gene expression in samples derived
XX CC from human placenta. The probes are useful for antenatal diagnosis of
XX CC human genetic disorders.
XX XX
XX SQ Sequence 81 AA;

Query Match      8.4%; Score 81; DB 22; Length 81;
Best Local Similarity 100.0%; Pred. No. 2.7e-73;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 578 RFINDMWKIYSITATNAVDGVASSCRACALGSGSGSCVPCPPGHHYIEKETNCKECP 637
DB 1 RFINDMWKIYSITATNAVDGVASSCRACALGSGSGSCVPCPPGHHYIEKETNCKECP 60

QY 638 DTYLSTHGVYKREACIPCGPG 658
DB 61 DTYLSTHGVYKREACIPCGPG 81

RESULT 10
ABG43186
ID ABG43186 standard; Peptide; 81 AA.
AC ABG43186;
XX XX
XX DT 19-AUG-2002 (first entry)
XX XX
XX DE Human peptide encoded by genome-derived single exon probe SEQ ID 32851.
XX XX
XX KW Human; single exon probe; asthma; lung cancer; COPD; ILD;
XX KW chronic obstructive pulmonary disease; interstitial lung disease;
XX KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
XX KW pulmonary alveolar proteinosis; lymphangioloeyomatosis; Karagener syndrome;
XX KW primary ciliary dyskinesia; fibrocystic pulmonary dysplasia;
XX KW hyaline membrane disease.
XX OS Homo sapiens.
XX PN WO200186003-A2.
XX XX
XX PD 15-NOV-2001.
XX XX
XX PF 30-JAN-2001; 2001WO-US00665.
XX XX
XX PR 04-FEB-2000; 2000US-180312P.
XX PR 26-MAY-2000; 2000US-207456P.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-234687P.
XX PR 27-SEP-2000; 2000US-236359P.

```

AC AAB70281;
XX
XX DT 10-MAY-2001 (first entry)
XX DE XX
XX DE Peptide #25.
XX
KW TR16 recptor; tumour necrosis factor receptor superfamily;
KW apoptosis; inflammatory; cancer; immune; neurodegenerative.
XX OS
XX Unidentified.
XX
XX PN WO200112671-A1.
XX PD
XX 22-FEB-2001.
XX
XX PF 10-AUG-2000; 2000WO-US21895.
XX
XX PR 12-AUG-1999; 99US-0148348.
XX PR 13-AUG-1999; 99US-0148683.
XX PR 13-AUG-1999; 99US-0148870.
XX PR 16-AUG-1999; 99US-0148758.
XX PR 17-AUG-1999; 99US-0149181.
XX PR 18-AUG-1999; 99US-0149453.
XX PR 19-AUG-1999; 99US-0149496.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX PA
XX Ruben SM, Young PE, Baker KP;
XX PI
XX WPI; 2001-138754/14.
XX DR
XX New nucleic acid molecule encoding a TR16 tumor necrosis factor
XX PT receptor polypeptide, useful for the diagnosis and treatment of cancer,
XX PT autoimmune disorders and cardiovascular diseases -
XX
XX PS Disclosure; Page 81; 286pp; English.
XX
XX CC The present invention relates to a TR16 receptor (tumour necrosis
XX CC factor receptor superfamily). The invention is useful treating
XX CC diseases and disorders associated with the inhibited or increased
XX CC apoptosis. In particular inflammatory diseases, cancers, immune and
XX CC neurodegenerative disorders may be treated.
XX
XX SQ Sequence 74 AA;

Query Match 7.7%; Score 74; DB 22; Length 74;
Best Local Similarity 100.0%; Pred. No. 2.9e-66;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps

QY 338 CTERPCTTKDFVQHTPCDEGKTOIMKYKIEPKICREDLTAIRLPSPGEKKDCPPCN 397
Db 1 CTERPCTTKDFVQHTPCDEGKTOIMKYKIEPKICREDLTAIRLPSPGEKKDCPPCN 60
|||||
QY 398 PGFYNGSSSCHPC 411
Db 61 PGFYNGSSSCHPC 74
|||||

RESULT 12
AAM72925
ID ID AAM72925 standard; Protein; 64 AA.
XX
XX AC AAM72925;
XX
XX DT 06-NOV-2001 (first entry)
XX
XX Human bone marrow expressed probe encoded protein SEQ ID NO: 33231.
XX
XX Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX
XX Homo sapiens.
OS

PN WO200157276-A2.
 XX 09-AUG-2001.
 PD
 PF
 PP
 XX 30-JAN-2001; 2001WO-US00668.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-488900/53.
 DR
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow.
 PS Example 4; SEQ ID NO: 33231; 658pp + Sequence Listing; English.
 XX
 CC The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is a
 CC protein encoded by one of the probes of the invention.
 XX
 XX Sequence 64 AA;
 SQ
 Query Match 6.6%; Score 64; DB 22; Length 64;
 Best Local Similarity 100.0%; Pred. No. 3.2e-56;
 Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 105 FSCASGEYLEMKNQVCKGEGTSLGSGIKFDEWDELPAFGFSNIATFMDTVVGPDSRP 164
 DB 1 FSCASGEYLEMKNQVCKGEGTSLGSGIKFDEWDELPAFGFSNIATFMDTVVGPDSRP 60
 QY 165 DGCN 168
 DB 61 DGCN 64
 RESULT 13
 ABG42760
 ID ABG42760 standard; Peptide; 64 AA.
 XX
 AC ABG42760;
 XX
 DT 19-AUG-2002 (first entry)
 XX
 DE Human peptide encoded by genome-derived single exon probe SEQ ID 32425.
 XX
 KW Human; single exon probe; asthma; lung cancer; COPD; ILD;
 KW chronic obstructive pulmonary disease; interstitial lung disease;
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 KW primary ciliary dyskinesia; pulmonary hypertension;
 KW hyaline membrane disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200186003-A2.
 XX
 PD 15-NOV-2001.

PF 30-JAN-2001; 2001WO-US00665.
 XX 04-FEB-2000; 2000US-180312P.
 PR 26-MAY-2000; 2000US-207456P.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-234687P.
 PR 27-SEP-2000; 2000US-236359P.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2002-114183/15.
 DR
 XX Spatially-addressable set of single exon nucleic acid probes, used to
 PT measure gene expression in human lung samples -
 XX
 XX Claim 27; SEQ ID NO 32425; 634pp; English.
 PS
 XX The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of
 CC probes; the novel set of probes which hybridise at high stringency to a
 CC nucleic acid expressed in the human lung; measuring gene expression in a
 CC sample derived from human lung, comprising (a) contacting the array with
 CC a collection of detectably labeled nucleic acids derived from human lung
 CC mRNA, and (b) measuring the label detectably bound to each probe of
 CC the array; identifying exons in a eukaryotic genome, comprising
 CC (a) algorithmically predicting at least one exon from genomic sequences
 CC of labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridisation to a single exon
 CC microarrays having a probe with the exon, where a common pattern of
 CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences, mentioned in the specification, or encoded by the
 CC probes/open reading frames (ORF). The probes are used for gene
 CC expression analysis, and for identifying exons in a gene, particularly
 CC using human lung derived mRNA and for the study of lung diseases
 CC such as asthma, lung cancer, chronic obstructive pulmonary disease
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
 CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
 CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
 CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
 CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
 CC and hyaline membrane disease. The present sequence is a peptide/protein
 CC encoded by a single exon probe of the invention.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 64 AA;
 Query Match 6.6%; Score 64; DB 23; Length 64;
 Best Local Similarity 100.0%; Pred. No. 3.2e-56;
 Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 105 FSCASGEYLEMKNQVCKGEGTSLGSGIKFDEWDELPAFGFSNIATFMDTVVGPDSRP 164
 DB 1 FSCASGEYLEMKNQVCKGEGTSLGSGIKFDEWDELPAFGFSNIATFMDTVVGPDSRP 60
 QY 165 DGCN 168

XX KW Human; gene expression; heart; microarray; vascular system;
KW KW cardiovascular disease; hypertension; cardiac arrhythmia;
XX KW congenital heart disease.
OS Homo sapiens.
XX WO200157274-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00666.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WIPI; 2001-48899/53.
XX Single exon nucleic acid probes for analyzing gene expression in human
XX hearts -
XX PS Claim 15; SEQ ID NO 25529; 530pp; English.
XX CC The present invention relates to single exon nucleic acid probes for
XX CC measuring human gene expression in a sample derived from human heart (see
XX CC ABA21535-ABA41305). The present sequence is a protein encoded by one such
XX CC probe. The probes may be used for predicting, measuring and displaying
XX CC gene expression in samples derived from the human heart via microarrays.
XX CC By measuring gene expression, the probes are useful for predicting,
XX CC diagnosing, grading, staging, monitoring and prognosing diseases of the
XX CC human heart and vascular system e.g. cardiovascular disease,
XX CC hypertension, cardiac arrhythmias and congenital heart disease.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 60 AA;
Query Match 6.2%; Score 60; DB 22; Length 60;
Best Local Similarity 100.0%; Pred. No. 3.3e-52;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 362 TOIMYKWIPEKICREDLTDALRLPPSGEKKDCPPCNPFGYNNGSSSCHPCPPGTFSDGK 421
DB 1 TOIMYKWIPEKICREDLTDALRLPPSGEKKDCPPCNPFGYNNGSSSCHPCPPGTFSDGK 60
Search completed: May 12, 2003, 13:38:52
Job time : 35.8744 secs

Db 61 DGCN 64
||||
RESULT 14
ABB38686
ID ABB38686 standard; Peptide; 60 AA.
XX AC ABB38686;
XX DF 04-FEB-2002 (first entry)
XX DE Peptide #6192 encoded by human foetal liver single exon probe.
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX OS Homo sapiens.
XX PN WO200157277-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00669.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WIPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human foetal liver -
XX PS Claim 27; SEQ ID NO 31321; 639pp + sequence listing; English.
XX CC The invention relates to a single exon nucleic acid probe for
XX CC measuring human gene expression in a sample derived from human foetal
XX CC liver. The single exon nucleic acid probes may be used for predicting,
XX CC measuring and displaying gene expression in samples derived from human
XX CC foetal liver. The present sequence is a peptide encoded by a single exon
XX CC nucleic acid probe of the invention.
XX CC Note: The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from wipo.int/pub/published_pct_sequences.
XX SQ Sequence 60 AA;
Query Match 6.2%; Score 60; DB 22; Length 60;
Best Local Similarity 100.0%; Pred. No. 3.3e-52;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 362 TOIMYKWIPEKICREDLTDALRLPPSGEKKDCPPCNPFGYNNGSSSCHPCPPGTFSDGK 421
DB 1 TOIMYKWIPEKICREDLTDALRLPPSGEKKDCPPCNPFGYNNGSSSCHPCPPGTFSDGK 60
RESULT 15
ABB23759
ID ABB23759 standard; Protein; 60 AA.
XX AC ABB23759;
XX DF 23-JAN-2002 (first entry)
XX DE Protein #5758 encoded by probe for measuring heart cell gene expression.

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run On: May 12, 2003, 13:31:24 ; Search time 16.4533 Seconds
(without alignments)
5626.691 Million cell updates/sec

Title: US-10-073-333A-2
Perfect score: 963
Sequence: 1 MLFRARGFVRGWRGPAEA.....TCYFWKKNQKKKTLNLFN 963

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_73.*

1: p1r1.*

2: p1r2.*

3: p1r3.*

4: p1r4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	0.8	100	1	defensin alpha-6 p
2	8	0.8	148	2	hypothetical prote
3	8	0.8	184	1	BLI protein - toma
4	8	0.8	228	2	S65426
5	8	0.8	228	2	pyruvate decarboxy
6	8	0.8	239	2	conserved hypothet
7	8	0.8	268	2	E. coli RpiR trans
8	8	0.8	268	2	AH1797
9	8	0.8	271	2	OX40 antigen precu
10	8	0.8	272	2	gene ox40 protein
11	8	0.8	281	2	hypothetical prote
12	8	0.8	286	2	ATP synthase gamma
13	8	0.8	405	2	S65471
14	8	0.8	417	2	S57820
15	8	0.8	428	2	pyruvate decarboxy
16	8	0.8	444	2	probable coenzyme
17	8	0.8	463	2	protein kinase hom
18	8	0.8	585	2	hypothetical prote
19	8	0.8	603	2	pyruvate decarboxy
20	8	0.8	622	2	pyruvate decarboxy
21	8	0.8	671	2	fructanase - Bacte
22	8	0.8	742	2	kinesin-related pr
23	8	0.8	881	2	hypothetical prote
24	8	0.8	1206	2	serine/threonine-s
25	8	0.8	1578	2	probable conserved
26	8	0.8	1582	2	peptidoglycan boun
27	8	0.8	2295	2	adhesin homolog lm
28	8	0.8	3102	2	probable membrane
29	7	0.7	94	2	laminin alpha chai
30	7	0.7	94	2	hypothetical prote

30 7 0.7 100 1 TNLJSI
31 7 0.7 100 2 A43998
32 7 0.7 109 2 S42599
33 7 0.7 122 2 T28199
34 7 0.7 134 2 C84023
35 7 0.7 134 2 T36365
36 7 0.7 135 2 PC2269
37 7 0.7 141 2 A27482
38 7 0.7 147 2 E70746
39 7 0.7 157 2 AG3560
40 7 0.7 159 2 B29879
41 7 0.7 173 2 AB3450
42 7 0.7 179 2 H64472
43 7 0.7 180 2 T49530
44 7 0.7 182 2 A64834
45 7 0.7 182 2 C85620

ALIGNMENTS

RESULT 1
S27016
defensin alpha-6 precursor - human
N:Alternate names: Paneth cell-specific alpha-defensin 6
C:Species: Homo sapiens (man)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: S27016
R:Jones, D.E.; Bevins, C.L.
FEBS Lett. 315, 187-192, 1993
A:Title: Defensin-6 mRNA in human Paneth cells: implications for antimicrobial peptid
A:Reference number: S27016; MUID:93114459; PMID:841777
A:Accession: S27016
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-100 <JON>
A:Cross-references: GB:M98331; MID:g181546; PIDN:AAB59357.1; PID:g181547
C:Genetics:
A:Gene: GDB:DEFA6; DEF6; HD-6
A:Cross-references: GDB:I36838; OMIM:600471
A:Map position: 8pter-8p21
C:Superfamily: mammalian defensin
C:Keywords: antibiotic; homodimer
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-70/Domain: propeptide #status predicted <PRO>
F:71-100/Product: defensin alpha-6 #status predicted <MAT>
F:72-99,74-88,78-98/Disulfide bonds: #status predicted

Query Match 0.8%; Score 8; DB 1; Length 100;
Best Local Similarity 100.0%; Pred. No. 4.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 936 TAVLLVAL 943
Db 8 TAVLLVAL 15

RESULT 2
E84239
hypothetical protein Vng0825c [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: E84239
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.;
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: E84239
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-148 <STO>
A:Cross-references: GB:AE004437; NID:gl0580395; PIDN:AAG19281.1; GSPDB:GN00138
C:Genetics:
A:Gene: VNG0825C

Query Match 0.8%; Score 8; DB 2; Length 148;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 155 TVGSPDS 162
|||||
Db 5 TVGSPDS 12

RESULT 3

QOCVLG

BL1 protein - tomato golden mosaic virus
C:Species: tomato golden mosaic virus
A:Note: host Nicotiana sp. (tobacco)
C:Date: 28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change 08-Apr-1994
C:Accession: A04169
R:Hamilton, W.D.O.; Stein, V.E.; Coutts, R.H.A.; Buck, K.W.
EMBO J. 3, 2197-2205, 1984
A:Title: Complete nucleotide sequence of the infectious cloned DNA components of tomato
A:Reference number: A04163
A:Accession: A04169
A:Molecule type: DNA
A:Residues: 1-184 <HAM>
C:Comment: The genome consists of two circular, single-stranded DNA components, DNA A and
C:Genetics:
A:Map position: segment B
C:Superfamily: tomato golden mosaic virus BL1 protein

Query Match 0.8%; Score 8; DB 1; Length 184;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 177 NYIESNRD 184
|||||
Db 13 NYIESNRD 20

RESULT 4

S65426
pyruvate decarboxylase (EC 4.1.1.1) - fava bean (fragment)
C:Species: Vicia faba (fava bean)
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 15-Oct-1999
C:Accession: S65426
R:Muecke, U.; Wohlfarth, T.; Fiedler, U.; Baumelein, H.; Ruecknagel, K.P.; Koenig, S.
Eur. J. Biochem. 237, 373-382, 1996
A:Title: Pyruvate decarboxylase from Pisum sativum. Properties, nucleotide and amino acid
A:Reference number: S65423; MUID:96215432; PMID:8647075
A:Accession: S65426
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-228 <MUE>
C:Superfamily: acetolactate synthase large chain; thiamin pyrophosphate-binding domain
C:Keywords: carbon-carbon lyase; carboxy-lyase

Query Match 0.8%; Score 8; DB 2; Length 228;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 269 KAVKPVLV 276
|||||
Db 50 KAVKPVLV 57

RESULT 5

S65425
pyruvate decarboxylase (EC 4.1.1.1) (clone PDC3) - garden pea (fragment)
C:Species: Pisum sativum (garden pea)

C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 15-Oct-1999
C:Accession: S65425
R:Muecke, U.; Wohlfarth, T.; Fiedler, U.; Baumelein, H.; Ruecknagel, K.P.; Koenig, S.
Eur. J. Biochem. 237, 373-382, 1996
A:Title: Pyruvate decarboxylase from Pisum sativum. Properties, nucleotide and amino
A:Reference number: S65423; MUID:96215432; PMID:8647075
A:Accession: S65425
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-228 <MUE>
C:Superfamily: acetolactate synthase large chain; thiamin pyrophosphate-binding domain
C:Keywords: carbon-carbon lyase; carboxy-lyase

Query Match 0.8%; Score 8; DB 2; Length 228;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 269 KAVKPVLV 276
|||||
Db 50 KAVKPVLV 57

RESULT 6

G87265
conserved hypothetical protein CC0136 [Imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: G87265
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-239 <STO>
A:Cross-references: GB:AE005673; NID:gl3421247; PIDN:AAK22123.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC0136

Query Match 0.8%; Score 8; DB 2; Length 239;
Best Local Similarity 100.0%; Pred. No. 9.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 AEAPRRGR 25
|||||
Db 176 AEAPRRGR 183

RESULT 7

AB1424
E. coli RpiR transcription regulator homolog lmo2795 [Imported] - Listeria monocytogenes
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi,
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkut, G.; Madueno, E.; Maltournam, A.;
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
A:Title: Comparative-genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AB1424
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-268 <GLA>
A:Cross-references: GB:NC_003210; PIDN:CAD01008.1; PID:gl6412295; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo2795

```
Query Match      0.8%; Score 8; DB 2; Length 268;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 792 LKNINIKE 799
Db 32 LKNINIKE 39

RESULT 8
AHL197
E. coli Rpr transcription regulator homolog lin2927 [imported] - Listeria innocua (strain
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AHL197
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AHL197
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-268 <GLA>
A:Cross-references: GB:AL592022; PIDN:CA98152.1; PID:q16415468; GSPDB:GN00178
A:Experimental source: strain Clp11262
C:Genetics:
A:Gene: lin2927

Query Match      0.8%; Score 8; DB 2; Length 268;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 792 LKNINIKE 799
Db 32 LKNINIKE 39

RESULT 9
SL2783
OX40 antigen precursor - rat
N:Alternate names: nerve growth factor receptor homolog
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 05-Nov-1999
C:Accession: SL2783; S08036
R:Maliet, S.; Fossum, S.; Barclay, A.N.
EMBO J. 9, 1063-1068, 1990
A:Title: Characterization of the MRC OX40 antigen of activated CD4 positive T lymphocyte
A:Reference number: SL2783; MUID:90214614; PMID:2157591
A:Accession: SL2783
A:Molecule type: mRNA
A:Residues: 1-271 <MAL>
A:Cross-references: EMBL:X17037; NID:g57830; PIDN:CAA34897.1; PID:g57831
C:Superfamily: CD27 antigen; NGF receptor repeat homology
C:Keywords: growth factor receptor; transmembrane protein
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-27/Product: OX40 antigen #status predicted <NAT>
F:211-235/Domain: transmembrane #status predicted <TM>

Query Match      0.8%; Score 8; DB 2; Length 271;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCCPGH 623
Db 122 CVPCCPGH 129

RESULT 10
AHL197
E. coli Rpr transcription regulator homolog lin2927 [imported] - Listeria innocua (stra
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AHL197
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AHL197
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-268 <GLA>
A:Cross-references: GB:AL592022; PIDN:CA98152.1; PID:q16415468; GSPDB:GN00178
A:Experimental source: strain Clp11262
C:Genetics:
A:Gene: lin2927

Query Match      0.8%; Score 8; DB 2; Length 268;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 792 LKNINIKE 799
Db 32 LKNINIKE 39

RESULT 11
T33466
hypothetical protein C43H8.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T33466
R:Tin-Wollam, A.M.
submitted to the EMBL Data Library, October 1998
A:Description: The sequence of C. elegans cosmid C43H8.
A:Reference number: 221350
A:Accession: T33466
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-281 <TIN>
A:Cross-references: EMBL:AF098499; PIDN:AAC67398.1; GSPDB:GN00019; CESP:C43H8.3
A:Experimental source: strain Bristol N2; clone C43H8
C:Genetics:
A:Gene: CESP:C43H8.3
A:Map position: 1
A:Introns: 58/1; 114/2; 242/3

Query Match      0.8%; Score 8; DB 2; Length 281;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 NITDFTVK 738
Db 111 NITDFTVK 118

RESULT 12
A82929
ATP synthase gamma chain UUI30 [imported] - Ureaplasma urealyticum
C:Species: Ureaplasma urealyticum
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
```

C:Accession: A82929
R:Glass, J.L.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
submitted to GenBank, February 2000
A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a min
A:Reference number: A82870
A:Accession: A82929
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-286 <GLA>
A:Cross-references: GB:AB002114; GB:AF222894; NID:g6899086; PIDN:AAF30536.1; GSPDB:GN001
A:Experimental source: serovar 3; biovar 1
C:Genetics:
A:Gene: atpC; U0130
A:Genetic code: SGC3
C:Superfamily: H+-transporting ATP synthase gamma chain

Query Match 0.8%; Score 8; DB 2; Length 286;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 739 EIVAGSDD 746
|||||
Db 279 EIVAGSDD 286

RESULT 13
S65471
pyruvate decarboxylase (EC 4.1.1.1) (clone PDC2) - garden pea (fragment)
C:Species: Pisum sativum (garden pea)
C:Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 20-Jun-2000
C:Accession: S65471; S65424
R:Muecke, U.; Wohlfarth, T.; Fiedler, U.; Baeumlein, H.; Ruecknagel, K.P.; Koenig, S.
Eur. J. Biochem. 237, 373-382, 1996
A:Title: Pyruvate decarboxylase from Pisum sativum. Properties, nucleotide and amino aci
A:Reference number: S65423; MUID:96215432; PMID:8647075
A:Accession: S65471
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-405 <MUE>
A:Cross-references: EMBL:Z66544; NID:g1177604; PIDN:CAA91445.1; PID:g1177605
A:Accession: S65424
A:Molecule type: protein
A:Residues: 1-8,146-153;209-218;350-362;386-393 <MUW>
C:Genetics:
A:Gene: pdc
C:Superfamily: acetolactate synthase large chain; thiamin pyrophosphate-binding domain
C:Keywords: carbon-carbon lyase; carboxy-lyase
F:1-405/Product: pyruvate decarboxylase #status experimental <MAT>
F:271-317/Domain: thiamin pyrophosphate-binding domain homology <TPB>

Query Match 0.8%; Score 8; DB 2; Length 405;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 KAVKPVLV 276
|||||
Db 50 KAVKPVLV 57

RESULT 14
S57820
pyruvate decarboxylase (EC 4.1.1.1) 1 - common tobacco (fragment)
C:Species: Nicotiana tabacum (common tobacco)
C:Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 15-Oct-1999
C:Accession: S57820
R:Butcher, M.; Brander, K.A.; Sbicego, S.; Mandel, T.; Kuhlmeier, C.
Plant Mol. Biol. 28, 739-750, 1995
A:Title: Aerobic fermentation in tobacco pollen.
A:Reference number: S57819; MUID:95375236; PMID:7647304
A:Accession: S57820
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-417 <BUC>

A:Cross-references: EMBL:X81854
C:Superfamily: acetolactate synthase large chain; thiamin pyrophosphate-binding domai
C:Keywords: carbon-carbon lyase; carboxy-lyase

Query Match 0.8%; Score 8; DB 2; Length 417;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 KAVKPVLV 276
|||||
Db 155 KAVKPVLV 162

RESULT 15
AF0241
probable coenzyme synthetase YPO1981 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C:Accession: AF0241
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrel
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AF0241
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-428 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC90794.1; PID:g15979994; GSPDB:GN00175
C:Genetics:
A:Gene: YPO1981

Query Match 0.8%; Score 8; DB 2; Length 428;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 604 ACALGSEQ 611
|||||
Db 84 ACALGSEQ 91

Search completed: May 12, 2003, 13:41:38
Job time : 19.4533 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 13:37:44 ; Search time 16.5146 seconds
(without alignments)
1829.738 Million cell updates/sec

Title: US-10-073-333A-4
Perfect score: 1027
Sequence: 1 MLFRARGVRGRGWRPAEA.....KEKEDHFSVOLKTRSPNI 1027

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0
Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Listing first 45 summaries

Database : Issued Patents.AA.*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	0.9	1487	US-08-760-489-2	Sequence 2, Appli
2	9	0.9	1487	US-08-760-489-4	Sequence 4, Appli
3	9	0.9	1487	US-09-185-373-2	Sequence 2, Appli
4	9	0.9	1487	US-09-185-373-4	Sequence 4, Appli
5	8	0.8	100	US-08-158-189-7	Sequence 7, Appli
6	8	0.8	133	US-08-468-560C-8	Sequence 8, Appli
7	8	0.8	139	US-08-219-237B-8	Sequence 8, Appli
8	8	0.8	140	US-08-477-347-17	Sequence 17, Appli
9	8	0.8	140	US-08-476-862-8	Sequence 8, Appli
10	8	0.8	153	US-09-286-529-2	Sequence 2, Appli
11	8	0.8	205	US-08-974-022-51	Sequence 51, Appli
12	8	0.8	205	US-08-795-445A-51	Sequence 51, Appli
13	8	0.8	205	US-08-795-447A-51	Sequence 51, Appli
14	8	0.8	205	US-08-974-186-51	Sequence 51, Appli
15	8	0.8	205	US-08-795-446B-51	Sequence 51, Appli
16	8	0.8	205	US-08-706-945D-138	Sequence 138, App
17	8	0.8	206	US-08-097-827-7	Sequence 7, Appli
18	8	0.8	206	US-08-494-574-7	Sequence 7, Appli
19	8	0.8	211	US-09-286-529-20	Sequence 20, Appli
20	8	0.8	299	US-09-286-529-17	Sequence 17, Appli
21	8	0.8	300	US-08-794-796-2	Sequence 2, Appli
22	8	0.8	438	US-08-097-827-11	Sequence 11, Appli
23	8	0.8	438	US-08-494-574-11	Sequence 11, Appli
24	7	0.7	15	US-08-602-999A-363	Sequence 363, App
25	7	0.7	15	US-09-500-124-363	Sequence 25, App
26	7	0.7	77	US-09-146-950-25	Sequence 25, Appli
27	7	0.7	100	US-08-467-822-23	Sequence 23, Appli

28	7	0.7	100	4	US-08-432-697-23	Sequence 23, Appli
29	7	0.7	100	4	US-08-466-248-23	Sequence 23, Appli
30	7	0.7	155	4	US-09-146-950-4	Sequence 4, Appli
31	7	0.7	159	4	US-09-146-950-20	Sequence 20, Appli
32	7	0.7	193	4	US-09-146-950-2	Sequence 2, Appli
33	7	0.7	197	4	US-09-146-950-18	Sequence 18, Appli
34	7	0.7	278	2	US-08-569-168-2	Sequence 2, Appli
35	7	0.7	283	4	US-08-509-024-2	Sequence 2, Appli
36	7	0.7	283	4	US-09-333-279-2	Sequence 2, Appli
37	7	0.7	283	4	US-09-072-993C-2	Sequence 2, Appli
38	7	0.7	283	5	PCT-US96-12374-2	Sequence 2, Appli
39	7	0.7	320	2	US-08-933-750C-12	Sequence 12, Appli
40	7	0.7	320	4	US-09-234-613-12	Sequence 12, Appli
41	7	0.7	331	4	US-09-086-483A-3	Sequence 3, Appli
42	7	0.7	383	2	US-08-569-168-7	Sequence 7, Appli
43	7	0.7	419	4	US-08-509-024-7	Sequence 7, Appli
44	7	0.7	419	4	US-09-333-279-7	Sequence 7, Appli
45	7	0.7	479	4	US-09-177-349-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-760-489-2
; Sequence 2, Application US/08760489
; Patent No. 5830696
; GENERAL INFORMATION:
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: DIRECTED EVOLUTION OF THERMOPHILIC
; TITLE OF INVENTION: ENZYMES
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/760,489
FILING DATE: 05-DEC-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
PRIOR APPLICATION NUMBER: 60/008,311
FILING DATE: 07-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Haile, Ph.D., Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/008001
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1487 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-760-489-2

Query Match 0.9% Score 9; DB 2; Length 1487;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 19 EAPRRGRSP 27
|||||

```

; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/185,373
; FILING DATE: 03-NO. 6335179-1998
; CLASSIFICATION: <Unknown>
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US/08/760,489
; FILING DATE: 05-DEC-1996
; APPLICATION NUMBER: 60/008,311
; FILING DATE: 07-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/008001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1487 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-185-373-2
Query Match 0.9%; Score 9; DB 4; Length 1487;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 9; Conservative 0; Mismatches 0; Indels
Qy 19 EAPRRGRSP 27
Db 171 EAPRRGRSP 179
RESULT 4
US-09-185-373-4
; Sequence 4, Application US/09185373
; Patent No. 6335179
; GENERAL INFORMATION:
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: DIRECTED EVOLUTION OF THERMOPHILIC
; ENZYMES
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/185,373
; FILING DATE: 03-NO. 6335179-1998
; CLASSIFICATION: <Unknown>
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US/08/760,489
; FILING DATE: 05-DEC-1996
; APPLICATION NUMBER: 60/008,311
; FILING DATE: 07-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/008001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1487 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-185-373-2
Query Match 0.9%; Score 9; DB 2; Length 1487;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 9; Conservative 0; Mismatches 0; Indels
Qy 19 EAPRRGRSP 27
Db 171 EAPRRGRSP 179
RESULT 3
US-09-185-373-2
; Sequence 2, Application US/09185373
; Patent No. 6335179
; GENERAL INFORMATION:
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: DIRECTED EVOLUTION OF THERMOPHILIC
; ENZYMES
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:

```

TELECOMMUNICATION INFORMATION:
 TELEPHONE: 619/678-5070
 TELEFAX: 619/678-5099
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1487 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 FRAGMENT TYPE: internal
 SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Query Match 0.9%; Score 9; DB 4; Length 1487;
 Best Local Similarity 100.0%; Pred. No. 3.2;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 EAPRRGRSP 27
 DB 171 EAPRRGRSP 179

RESULT 5
 US-08-158-189-7
 ; Sequence 7, Application US/08158189
 ; Patent No. 5641497
 ; GENERAL INFORMATION:
 ; APPLICANT: Bevis, Charles L.
 ; APPLICANT: Jones, Douglas E.
 ; TITLE OF INVENTION: Gastrointestinal Defensin Peptides,
 ; NUMBER OF SEQUENCES: 51
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz and No. 5641497ris
 ; STREET: One Liberty Place - 46th Floor
 ; CITY: Philadelphia
 ; STATE: PA
 ; COUNTRY: U.S.A.
 ; ZIP: 19103
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/158,189
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/888,232
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Johnson, Philip S.
 ; REGISTRATION NUMBER: 27,200
 ; REFERENCE/DOCKET NUMBER: CH-0219
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 215-568-3100
 ; TELEFAX: 215-568-3439
 ; INFORMATION FOR SEQ ID NO: 7:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 100 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-158-189-7

Query Match 0.8%; Score 8; DB 1; Length 100;
 Best Local Similarity 100.0%; Pred. No. 2.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 936 TAVLLVAL 943
 DB 8 TAVLLVAL 15

RESULT 6
 US-08-468-560C-8
 ; Sequence 8, Application US/08468560C
 ; Patent No. 6270998
 ; GENERAL INFORMATION:
 ; APPLICANT: NAGATA, Shigekazu
 ; APPLICANT: ITOH, Naoto
 ; APPLICANT: YONEHARA, Shin
 ; TITLE OF INVENTION: DNA CODING FOR HUMAN CELL SURFACE
 ; NUMBER OF SEQUENCES: 11
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP.
 ; STREET: P.O. BOX 747
 ; CITY: FALLS CHURCH
 ; STATE: VA
 ; COUNTRY: USA
 ; ZIP: 22040-0747
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/468,560C
 ; FILING DATE: 06-JUN-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: MURPHY JR., GERALD M.
 ; REGISTRATION NUMBER: 28,977
 ; REFERENCE/DOCKET NUMBER: 20-4393P
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 703-205-8000
 ; TELEFAX: 703-205-8050
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 133 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-468-560C-8

Query Match 0.8%; Score 8; DB 4; Length 133;
 Best Local Similarity 100.0%; Pred. No. 3.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCPPGH 623
 DB 91 CVPCPPGH 98

RESULT 7
 US-08-219-237B-8
 ; Sequence 8, Application US/08219237B
 ; Patent No. 5874546
 ; GENERAL INFORMATION:
 ; APPLICANT: NAGATA, Shigekazu
 ; APPLICANT: ITOH, Naoto
 ; APPLICANT: YONEHARA, Shin
 ; TITLE OF INVENTION: DNA Coding for Human Cell Surface Antigen
 ; NUMBER OF SEQUENCES: 11
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: James W. Hellwege
 ; STREET: P.O. Box 2266 Eads Station
 ; CITY: Arlington
 ; STATE: Virginia
 ; COUNTRY: USA
 ; ZIP: 22202
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/219,237B
FILING DATE: 28-MAR-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/872,129
FILING DATE: 22-APR-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: James W. Hellwege
REGISTRATION NUMBER: 28,808
REFERENCE/DOCKET NUMBER: 516762
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 139 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-219-237B-8

Query Match 0.8%; Score 8; DB 2; Length 139;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCCPPGH 623
DB 97 CVPCCPPGH 104

RESULT 8
US-08-477-347-17
Sequence 17, Application US/08477347
Patent No. 6232446
GENERAL INFORMATION:
APPLICANT: WALLACH, David
APPLICANT: BIGDA, Jacek
APPLICANT: BELETSKY, Igor
APPLICANT: METT, Igor
TITLE OF INVENTION: TNF LIGANDS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,347
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/115,685
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 106271
FILING DATE: 08-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Townsend, G. Kevin
REGISTRATION NUMBER: 34,033
REFERENCE/DOCKET NUMBER: WALLACH=10
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633

INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 140 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-477-347-17

Query Match 0.8%; Score 8; DB 4; Length 140;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCCPPGH 623
DB 98 CVPCCPPGH 105

RESULT 9
US-08-476-862-8
Sequence 8, Application US/08476862
Patent No. 6262239
GENERAL INFORMATION:
APPLICANT: WALLACH, David
APPLICANT: BIGDA, Jacek
APPLICANT: BELETSKY, Igor
APPLICANT: METT, Igor
APPLICANT: ENGELMANN, Hartmut
TITLE OF INVENTION: TNF INHIBITORS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/476,862
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 107267
FILING DATE: 12-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 94039
FILING DATE: 06-APR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 91229
FILING DATE: 06-AUG-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 90339
FILING DATE: 18-MAY-1989
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: WALLACH=12A
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 140 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-476-862-8

Query Match 0.8%; Score 8; DB 4; Length 140;
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCCPGH 623
DB 98 CVPCCPGH 105

RESULT 10
US-09-286-529-2
; Sequence 2, Application US/09286529
; Patent No. 6297367
; GENERAL INFORMATION:
; APPLICANT: Catherine Tribouley
; TITLE OF INVENTION: NEW MEMBERS OF TNF AND TNFR FAMILIES
; FILE REFERENCE: 1408.003/200130.439C1
; CURRENT APPLICATION NUMBER: US/09/286,529
; CURRENT FILING DATE: 1999-04-05
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 153
; TYPE: PRT
; ORGANISM: human
US-09-286-529-2

Query Match 0.8%; Score 8; DB 4; Length 153;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 410 PCPPGTFS 417
DB 68 PCPPGTFS 75

RESULT 11
US-08-974-022-51
; Sequence 51, Application US/08974022
; Patent No. 6015938
; GENERAL INFORMATION:
; APPLICANT: Boyle, William J.
; APPLICANT: Lacey, David L.
; APPLICANT: Calzone, Frank J.
; APPLICANT: Chang, Ming-Shi
; TITLE OF INVENTION: OSTEOPROTEGERIN
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/974,022
; FILING DATE: 12-DEC-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/577,788
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-378
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 205 amino acids

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-974-022-51
Query Match 0.8%; Score 8; DB 3; Length 205;
Best Local Similarity 100.0%; Pred. No. 5.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCCPGH 623
DB 122 CVPCCPGH 129
RESULT 12
US-08-795-445A-51
; Sequence 51, Application US/08795445A
; Patent No. 6284485
; GENERAL INFORMATION:
; APPLICANT: Boyle, William J.
; APPLICANT: Lacey, David L.
; APPLICANT: Calzone, Frank J.
; APPLICANT: Chang, Ming-Shi
; TITLE OF INVENTION: OSTEOPROTEGERIN
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/795,445A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/577,788
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-378
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 205 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-795-445A-51

Query Match 0.8%; Score 8; DB 4; Length 205;
Best Local Similarity 100.0%; Pred. No. 5.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCCPGH 623
DB 122 CVPCCPGH 129

RESULT 13
US-08-795-447A-51
; Sequence 51, Application US/08795447A
; Patent No. 6284728
; GENERAL INFORMATION:
; APPLICANT: Boyle, William J.
; APPLICANT: Lacey, David L.

APPLICANT: Calzone, Frank J.
APPLICANT: Chang, Ming-Shi
TITLE OF INVENTION: Osteoprotegerin
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen Inc.
STREET: One Amgen Center Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/975,447A
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Winter, Robert B.
REFERENCE/DOCKET NUMBER: A-378D2
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 205 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-795-447A-51

Query Match 0.8%; Score 8; DB 4; Length 205;
Best Local Similarity 100.0%; Pred. No. 5.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCCPGH 623
Db 122 CVPCCPGH 129

RESULT 14
US-08-974-186-51
Sequence 51, Application US/08974186
Patent No. 6284740
GENERAL INFORMATION:
APPLICANT: Boyle, William J.
APPLICANT: Lacey, David L.
APPLICANT: Calzone, Frank J.
APPLICANT: Chang, Ming-Shi
TITLE OF INVENTION: OSTEOPROTEGERIN
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen Inc.
STREET: 1840 Dehavilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,186
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/577,788
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Winter, Robert B.

REFERENCE/DOCKET NUMBER: A-378
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 205 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-974-186-51

Query Match 0.8%; Score 8; DB 4; Length 205;
Best Local Similarity 100.0%; Pred. No. 5.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCCPGH 623
Db 122 CVPCCPGH 129

RESULT 15
US-08-795-446B-51
Sequence 51, Application US/08795446B
Patent No. 6288032
GENERAL INFORMATION:
APPLICANT: Boyle, William J.
APPLICANT: Lacey, David L.
APPLICANT: Calzone, Frank J.
APPLICANT: Chang, Ming-Shi
TITLE OF INVENTION: OSTEOPROTEGERIN
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen Inc.
STREET: 1840 Dehavilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/795,446B
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/577,788
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Winter, Robert B.
REFERENCE/DOCKET NUMBER: A-378
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 205 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-795-446B-51

Query Match 0.8%; Score 8; DB 4; Length 205;
Best Local Similarity 100.0%; Pred. No. 5.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCCPGH 623
Db 122 CVPCCPGH 129

Search completed: May 12, 2003, 13:42:24
Job time : 18.5146 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 00:01:37 ; Search time 8811.67 Seconds
(without alignments)
11744.602 Million cell updates/sec

Title: US-10-073-333A-3
Perfect score: 3556
Sequence: 1 atgctgttcgcgcgcgggg.....ggtataaaaaaaaaaaaaa 3556

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries.

Database :

GenEmbl:*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg_hum.*
- 31: em.htg_inv.*
- 32: em.htg_other.*
- 33: em.htg_mus.*
- 34: em.htg_pln.*
- 35: em.htg_rod.*
- 36: em.htg_mam.*
- 37: em.htg_vrt.*
- 38: em.sy.*
- 39: em.htgo_hum.*
- 40: em.htgo_mus.*
- 41: em.htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2959.2	83.2	3185	9	AK055902	AK055902 Homo sapi
2	1499.6	42.2	1737	6	AX060311	AX060311 Sequence
3	1182.6	33.3	1508	6	AX060301	AX060301 Sequence
4	991.4	27.9	3280	6	AX036667	AX036667 Sequence
5	989.8	27.8	3331	6	AX127726	AX127726 Sequence
6	897.8	27.8	3501	6	AX463904	AX463904 Sequence
7	860.8	24.2	2610	6	AX127742	AX127742 Sequence
8	860.8	24.2	2733	6	AX127740	AX127740 Sequence
9	732.8	20.6	2895	9	AK057647	AK057647 Homo sapi
10	711.6	20.0	2188	10	BC029010	BC029010 Mus muscu
11	687	19.3	2449	9	BC031648	BC031648 Homo sapi
12	633.8	17.8	2129	9	AK026832	AK026832 Homo sapi
13	594	16.7	5567	9	AB037745	AB037745 Homo sapi
14	518.4	14.6	136150	9	AC002081	AC002081 Homo sapi
15	504	14.2	576	6	AX336856	AX336856 Sequence
16	494	13.9	1129	6	ARI84096	ARI84096 Sequence
17	487	13.7	1717	6	AX013071	AX013071 Sequence
18	403	11.3	404	6	AX071721	AX071721 Sequence
19	381.8	10.7	1149	6	AX127746	AX127746 Sequence
20	294.2	8.3	1149	10	BC022655	BC022655 Mus muscu
21	285.8	8.0	1119	6	AX213279	AX213279 Sequence
22	285.8	8.0	1587	6	AX213277	AX213277 Sequence
23	265.4	7.5	101719	2	AL357495	AL357495 Homo sapi
24	234.6	6.6	627	6	AX127744	AX127744 Sequence
25	217	6.1	131321	2	AC080160	AC080160 Mus muscu
26	197.2	5.5	165865	2	AC079958	AC079958 Mus muscu
27	196.4	5.5	101719	2	AL357495	AL357495 Homo sapi
28	189	5.3	107889	9	AC004519	AC004519 Homo sapi
29	180.4	5.1	108063	2	AC116297	AC116297 Rattus no
30	169.2	4.8	677	6	AX036669	AX036669 Sequence
31	161.6	4.5	426	6	AX368203	AX368203 Sequence
32	140.6	4.0	437	6	AX182039	AX182039 Sequence
33	135.4	3.8	480	6	AX182004	AX182004 Sequence
34	130.4	3.7	29979	9	AC000122	AC000122 Homo sapi
35	127.2	3.6	131321	2	AC080160	AC080160 Mus muscu
36	126.4	3.6	466	6	AX261103	AX261103 Sequence
37	123.6	3.5	353	6	AX182022	AX182022 Sequence
38	118.6	3.3	166384	2	AC113446	AC113446 Mus muscu
39	118	3.3	108063	2	AC116297	AC116297 Rattus no
40	115.6	3.2	150129	2	AC119789	AC119789 Rattus no
41	115	3.2	399	6	AX182000	AX182000 Sequence
42	114	3.2	183815	2	AC09081	AC09081 Rattus no
43	104	2.9	443	6	AX127734	AX127734 Sequence
44	99.8	2.8	97114	9	AL356389	AL356389 Human DNA
45	94	2.6	265	6	AX368154	AX368154 Sequence

ALIGNMENTS

RESULT 1

AK055902

LOCUS

DEFINITION

AK055902

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

AK055902 3185 bp mRNA linear PRI 01-AUG-2002
Homo sapiens cDNA FLJ31340 fis, clone MESAN1000035, weakly similar
to MAJOR SURFACE-LABELLED TROPHOZOITE ANTIGEN PRECURSOR.

AK055902.1 GI:16550745

oligo capping: fis (full insert sequence).

Homo sapiens normal mesangial cells (NHMC56046-2) cDNA to mRNA,

clone lib:MESAN1 clone:MESAN1000035.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1

Tashiro,H., Yamazaki,M., Watanabe,K., Kumagai,A., Itakura,S.,

Fukuzumi, Y., Fujimori, Y., Komiyama, M., Sugiyama, T., Irle, R.,
Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y.,
Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H.,
Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K.,
Wagatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A.,
Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S.,
Nagahari, K., Masuno, Y., Nagai, K. and Isogai, T.
NEDO human cDNA sequencing project

TITLE JOURNAL REFERENCE AUTHORS TITLE JOURNAL

Unpublished
2 (bases 1 to 3185)
Isogai, T., Otsuki, T. and Sugiyama, T.
Direct Submission
Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
Genomics Laboratory: 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan: cDNA full insert sequencing:
Research Association for Biotechnology (RAB): cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: RAB and
HRI.

FEATURES

source

Location/Qualifiers

1..3185
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MESAN100035"
/cell_type="normal mesangial cells (NHMC56046-2)"
/clone_lib="MESAN1"
/note="Cloning vector: pME18SFL3-primary culture, normal
mesangial cells"

CDS

362..1843
/note="unnamed protein product"
/codon_start=1
/db_xref="GI:16550746"
/translation="MLENGYVMLKSGTNILYWRITGLMGSKAVKPLVKNITIEGV
AYTSECFPKPGTSEKPGSPFNQVPCNTYSEKAKEICRCKDDSOFSSEGSSECTE
RPPCTDYFHITPCDEEGTQIMYKWIPIKREDLTDALRPLSPGEKKDCPPCNF
FYNGNSDQKPCPPGTFSDGTEKPCPCAGTEPALGFYKWNVLPGNMKTSQFNVG
NSKCDGMGWEVAGDIQSGAGGSDNDYLILNHPGPKPTSMGTSGELGRTFV
FETLCSADCVLYFMDVIRKSTNVVSMGGTKEOAVTHIIFKNATFTFTWAFORTNO
GQDNRRFLNDMVKLYSITATNAVGVASSCRALGSEQSSSCVPCPPGHYIKETN
CKRCPPDTYLSIHQVYKKEACIPCGPSKNNQHSVCDFYFPEKENOSLHYDFS
NLSVSGMLNGPSTSKGTQYFHEFNISYVGMRRRLSVPTI"
BASE COUNT 960 a 660 c 719 g 846 t

ORIGIN

Query Match 83.2%; Score 2959.2; DB 9; Length 3185;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 3032; Conservative 0; Mismatches 3; Indels 41; Gaps 3;

QY 504 CAACCTCTTCTGGATCCCTCGTGGAACTACATAAGATCTAATCGTGACTGCACCGT 563
DB 111 CAGCTCTTCTGGATCCCTCGTGGAACTACATAAGATCTAATCGTGACTGCACCGT 170
QY 564 GTCTTTGATCTATGCTGTCACCTTAAGAAGTCAGGCTATGCTCTTTTGATCAGTA 623
DB 171 GTCTTTGATCTATGCTGTCACCTTAAGAAGTCAGGCTATGCTCTTTTGATCAGTA 230
QY 624 TGTGCAACAACACTCTCTTTGAGTCTCTTTATTCAAATGATGATGCGCAGGAGATGA 683
DB 231 TGTGCAACAACACTCTCTTTGAGTCTCTTTATTCAAATGATGATGCGCAGGAGATGA 290
QY 684 CACCACCACTGACAAGTGGGTAAAACTTACAGACAATGGAGAATGGGGCTCTCATTTCT-- 741
DB 291 CACCACCACTGACAAGTGGGTAAAACTTACAGACAATGGAGAATGGGGCTCTCATTTCTGA 350
QY 742 -----GTAACTGCTGAATCAGGCACAAACATAC 769
DB 351 CAGCTCCTTAAATGCTTGAAGCAATTAACAGGTATGCTGAAATCAGGCACAAACATAC 410

QY 770 TCTACTGGAGAACTACAGGCATCTTATATGGGTTCTAAGCGGTCAAGCTGTGCTGGTAA 829
DB 411 TCTACTGGAGAACTACAGGCATCTTATATGGGTTCTAAGCGGTCAAGCTGTGCTGGTAA 470
QY 830 AAAATATCACATTTCAAGGGTGGGTACACATCAGATGTTTCTTTCGCAAGCCAGGCA 889
DB 471 AAAATATCACATTTCAAGGGTGGGTACACATCAGATGTTTCTTTCGCAAGCCAGGCA 530
QY 890 CATTTCAGCAACAACACAGGTTTCATTTCACTGCCAGGTGTGCCCAAGAACACTATTTCTG 949
DB 531 CATTTCAGCAACAACACAGGTTTCATTTCACTGCCAGGTGTGCCCAAGAACACTATTTCTG 590
QY 950 AGAAGAGCCCAAGAAATGTATANGGTGTAAAGACGACTCTCAATTTTC-----AGGAT 1003
DB 591 AGAAGAGCCCAAGAAATGTATANGGTGTAAAGACGACTCTCTCAATTTTCAGAGGAAGGAT 650
QY 1004 CCAGTGAGTGTACAGAGCCCTCCCTGTACCACAAAAGACTATTTCCAGATCCCATCTC 1063
DB 651 CCAGTGAGTGTACAGAGCCCTCCCTGTACCACAAAAGACTATTTCCAGATCCCATCTC 710
QY 1064 CATGTGATGAAGCAAGAACACACAGATAATGTACAAGTGGATAGAGCCCAAAATCTGCC 1123
DB 711 CATGTGATGAAGCAAGAACACACAGATAATGTACAAGTGGATAGAGCCCAAAATCTGCC 770
QY 1124 GGGAGGATCTCACAGATGCTATTAGATTGCCCTTCTGGAGAGAAGGATTTGTCGC 1183
DB 771 GGGAGGATCTCACAGATGCTATTAGATTGCCCTTCTGGAGAGAAGGATTTGTCGC 830
QY 1184 CTTGCAACCTCGGATTTTATACAAATGGATCATCTCTTGCCATCCCTGCTCCTGGAA 1243
DB 831 CTTGCAACCTCGGATTTTATACAAATGGATCATCTCTTGCCNTCCCTGCTCCTGGAA 890
QY 1244 CATTTTCAGATGGAAACCAAGAAATGTAGACCATGTCCAGCAGGAACGGAGCTGCACATTG 1303
DB 891 CATTTTCAGATGGAAACCAAGAAATGTAGACCATGTCCAGCAGGAACGGAGCTGCACATTG 950
QY 1304 GCTTTGATATATAATGGTGGAAATGTCCTTCTGCGCAACATGAAACTTCTCGCTTCAATG 1363
DB 951 GCTTTGATATATAATGGTGGAAATGTCCTTCTGCGCAACATGAAACTTCTCGCTTCAATG 1010
QY 1364 TTGGCAATTCAAAGTGGATGGAATGAATGGTGGAGGTGGCTGAGATCATATCCAGA 1423
DB 1011 TTGGCAATTCAAAGTGGATGGAATGAATGGTGGAGGTGGCTGAGATCATATCCAGA 1070
QY 1424 GTGGGGCTGGAGGTCTTGACAAATGATTAACCTTGATCTTAACTTGCATATCCAGAGTTTA 1483
DB 1071 GTGGGGCTGGAGGTCTTGACAAATGATTAACCTTGATCTTAACTTGCATATCCAGAGTTTA 1130
QY 1484 TACCACCAACATCTATGACTGAGCCACGGTTCCTGAACTAGGAAGAAATTAACATTTGCT 1543
DB 1131 TACCACCAACATCTATGACTGAGCCACGGTTCCTGAACTAGGAAGAAATTAACATTTGCT 1190
QY 1544 TTGAGACCTCTGTTTCAGCTGACTGTGTTTGTACTTCTTCTGATATTAATAGAAAA 1603
DB 1191 TTGAGACCTCTGTTTCAGCTGACTGTGTTTGTACTTCTTCTGATATTAATAGAAAA 1250
QY 1604 GTACAAATGTGGTAGAATCGTGGGTGGAAACAAAGAAAACAAAGCTTACACCCATATCA 1663
DB 1251 GTACAAATGTGGTAGAATCGTGGGTGGAAACAAAGAAAACAAAGCTTACACCCATATCA 1310
QY 1664 TCTTCAAGAATGCAACTTTTACATTTACATGGCAATTCAGAGAACTAATCAGGCTCAAG 1723
DB 1311 TCTTCAAGAATGCAACTTTTACATTTACATGGCAATTCAGAGAACTAATCAGGCTCAAG 1370
QY 1724 ATAAATAGAGGTTTCATCAATGACATGGTGAAGATTTTATTTCTATCAGCCACTAATCGAG 1783
DB 1371 ATAAATAGAGGTTTCATCAATGACATGGTGAAGATTTTATTTCTATCAGCCACTAATCGAG 1430
QY 1784 TTGATGGGGTGGCGTCTCATGCGCTGCTGCGCTCGGTTCTGCAACAGTGGGTTTCA 1843
DB 1431 TTGATGGGGTGGCGTCTCATGCGCTGCTGCGCTCGGTTCTGCAACAGTGGGTTTCA 1490
QY 1844 CGTGTGTCCTCCCTGCCCTCCAGGCCACTACATTGAGAAAAGAAACCAACCAAGTGCAGGAAT 1903

|||||
Db 1491 CGTGTGTCCTCCCTCCAGGCACTACATTGAGAAGAAACCAACAGTGAAGGAAT 1550
QY 1904 GTCCACCTGACACCTACCTGCTCCATACATCAGGTCTATGTCGAAGAGGCTTGATTCAT 1963
Db 1551 GTCCACCTGACACCTACCTGCTCCATACATCAGGTCTATGTCGAAGAGGCTTGATTCAT 1610
QY 1964 GCGGGCTGGGATGAAACAAATCAGGACCAATTCGGTTTCTATAGTACTGCTTTTCT 2023
Db 1611 GCGGGCTGGGATGAAACAAATCAGGACCAATTCGGTTTCTATAGTACTGCTTTTCT 1670
QY 2024 ACCATGAAGAAAGAAATCAGATTTTGCACCTATGACCTTACCACTCAGCAGTGGGCT 2083
Db 1671 ACCCTGAAAGAAAGAAATCAGATTTTGCACCTATGACCTTACCACTCAGCAGTGGGCT 1730
QY 2084 CATTATGAATGGCCCGCAGCTTCACCTCCAAAGGAACAAATATCTTCCATTTCTCAATA 2143
Db 1731 CATTATGAATGGCCCGCAGCTTCACCTCCAAAGGAACAAATATCTTCCATTTCTCAATA 1790
QY 2144 TCAGTTTATGTCGGCATGAGGGGAAGAGATGCTCTCTGTACCAACAAATATACAGACT 2203
Db 1791 TCAG-TTATGTCGGCATGAGGGGAAGAGATGCTCTCTGTACCAACAAATATACAGACT 1849
QY 2204 TTACAGTAAAGAAATAGTGGCAGGCTCAGATGATTACACAAATTTGGTAGGGGCATTTG 2263
Db 1850 TTACAGTAAAGAAATAGTGGCAGGCTCAGATGATTACACAAATTTGGTAGGGGCATTTG 1909
QY 2264 TATGCCAGTCAACAATATTCCTTCTGAAAGTAAGGTTTCCGAGCAGCCTTATCATCAC 2323
Db 1910 TATGCCAGTCAACAATATTCCTTCTGAAAGTAAGGTTTCCGAGCAGCCTTATCATCAC 1969
QY 2324 AATCCATCATCTGGCAGATACATTCATAGGAGTCAGAGTTGAAACCAACATGAAATA 2383
Db 1970 AATCCATCATCTGGCAGATACATTCATAGGAGTCAGAGTTGAAACCAACATGAAATA 2029
QY 2384 TTAATATAAAGAAATATGTTCCAGTTCCCAAGTCCCAACAGCCAAATACAGATGTCATTTCT 2443
Db 2030 TTAATATAAAGAAATATGTTCCAGTTCCCAAGTCCCAACAGCCAAATACAGATGTCATTTCT 2089
QY 2444 TTTATAAGTCTTCTACAGCAACAACATCTTGTATTATGCGCCATCAACTGCTGCGAAAA 2503
Db 2090 TTTATAAGTCTTCTACAGCAACAACATCTTGTATTATGCGCCATCAACTGCTGCGAAAA 2149
QY 2504 TGAGGTGTAATCTACTAATCTGAGCAGGAGTGAATTCAGTCCCGCAGCAAGTGCACG 2563
Db 2150 TGAGGTGTAATCTACTAATCTGAGCAGGAGTGAATTCAGTCCCGCAGCAAGTGCACG 2209
QY 2564 CAGGTACCTGTGATGGTGCTGCTTCTATTTCTGTCGGAGAGTCTGAAGCTTGCCCTC 2623
Db 2210 CAGGTACCTGTGATGGTGCTGCTTCTATTTCTGTCGGAGAGTCTGAAGCTTGCCCTC 2269
QY 2624 TGTGTACGGAGCATGACTTCCATGAGATTGAGGAGCGCTGCAAGAGAGATTTCAGGAAA 2683
Db 2270 TGTGTACGGAGCATGACTTCCATGAGATTGAGGAGCGCTGCAAGAGAGATTTCAGGAAA 2329
QY 2684 CCTGTGATGTGGAATGAACCTAATAGTGTCATTAAGGAATTTCTTCCCTGAGAAAA 2743
Db 2330 CCTGTGATGTGGAATGAACCTAATAGTGTCATTAAGGAATTTCTTCCCTGAGAAAA 2389
QY 2744 AGTTGCAACCTGTGAACGCTTGACTTTTGCTGAAGTGGGAGCGGCTGTTGGAGCTT 2803
Db 2390 AGTTGCAACCTGTGAACGCTTGACTTTTGCTGAAGTGGGAGCGGCTGTTGGAGCTT 2449
QY 2804 TTACTGCCGTTTTGCTGGTCTGACCTGCTTCTGTTGAAAAAGAAATCAAAAACCTGG 2863
Db 2450 TTACTGCCGTTTTGCTGGTCTGACCTGCTTCTGTTGAAAAAGAAATCAAAAACCTGG 2509
QY 2864 AATACAAATATTCAGATTAAGTAAAGCAGCTAATCAAAAGAGTGTGAACCTCCCGGCTG 2923
Db 2510 AATACAAATATTCAGATTAAGTAAAGCAGCTAATCAAAAGAGTGTGAACCTCCCGGCTG 2569
QY 2924 CAGACAGTTGCTATCATGGAAGGAGAGATAATCAAGAGAGTTGTATATTCAATA 2983
|||||



Db 2570 CAGACAGTTGCTGCTATCATGGAAGGAGAAGATAATGAAGAGGAAGTTGTATATTCCAATA 2629
QY 2984 AACAGTCACTACTAGGAAAACTCAATCTTTGGCAACCAAGGAAAAAGAACCATTTTG 3043
Db 2630 AACAGTCACTACTAGGAAAACTCAATCTTTGGCAACCAAGGAAAAAGAACCATTTTG 2689
QY 3044 AATCTGTTCAACTGAAAAACCTCAAGATCCCAATATATGAAGAGACAGTGTCTGACGCT 3103
Db 2690 AATCTGTTCAACTGAAAAACCTCAAGATCCCAATATATGAAGAGACAGTGTCTGACGCT 2749
QY 3104 TGAGACTAATGAACAAGAACCTGCTCTAGTTTACAGACCATATTTTAGGCTCTGTC 3163
Db 2750 TGAGACTAATGAACAAGAACCTGCTCTAGTTTACAGACCATATTTTAGGCTCTGTC 2809
QY 3164 CTCACTACCTGTCACATTTGTTGATCTCACAGAGGAGGCCATGCCGCTGAAAGGGAAGGA 3223
Db 2810 CTCACTACCTGTCACATTTGTTGATCTCACAGAGGAGGCCATGCCGCTGAAAGGGAAGGA 2869
QY 3224 GATTGAAACATTTGATTTGCCCTTATCACATGCTCAAGTACCTTGCCAAATAAAGGAAAGCA 3283
Db 2870 GATTGAAACATTTGATTTGCCCTTATCACATGCTCAAGTACCTTGCCAAATAAAGGAAAGCA 2929
QY 3284 AATGATTTGGGTCTCACTCACTCAAGTGAAGCTCAACTCAGGAAGAGATTTATCTGTATATA 3343
Db 2930 AATGATTTGGGTCTCACTCACTCAAGTGAAGCTCAACTCAGGAAGAGATTTATCTGTATATA 2989
QY 3344 CACATACTGAAACCAAGTTTAAGCCCACTGCTGCTGATGCCATGCCATATAATT 3403
Db 2990 CACATACTGAAACCAAGTTTAAGCCCACTGCTGCTGATGCCATGCCATATAATT 3049
QY 3404 AATGGTGAATTTTATTTATTTATGATGCTCTACATACTGATTTGGAAGGCACATG 3463
Db 3050 AATGGTGAATTTTATTTATTTATGATGCTCTACATACTGATTTGGAAGGCACATG 3109
QY 3464 TGAGCATATGCATATGATCCAAATTTATGTTTTTCTTTGTTTATATATTTTGGGAAAAAT 3523
Db 3110 TGAGCATATGCATATGATCCAAATTTATGTTTTTCTTTCTTTATATATTTTGGGAAAAAT 3169
QY 3524 AAAATTTTTTAAAGGT 3539
Db 3170 AAAATTTTTTAAAGGT 3185

RESULT 2
AX060311
LOCUS AX060311 1737 bp DNA linear PAT 22-JAN-2001
DEFINITION Sequence 19 from Patent WO0078802.
ACCESSION AX060311
VERSION AX060311.1 GI:12405800
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1737)
AUTHORS Shinkets, R.A., Fernandes, E., Vernet, C., Yang, M., Boldog, F.L. and Herrmann, J.L.
TITLE Secreted polypeptides and corresponding polynucleotides
JOURNAL Patent: WO 0078802-A 19 28-DEC-2000;
Curagen Corporation (US)
FEATURES
Location/Qualifiers
1..1737
/organism="Homo sapiens"
/db_xref="taxon:9606"
296..1690
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC24875.1"
/db_xref="GI:12405801"
/translation="MKQVCSKCGEGTSLGSGIKFDEWDELPAFSGNIATFMDTVWG
PDSRPDCGNSSWIPRGNYESNRDDCTSLIYAVHLKSGYVEFYQYVDNNIFFE
FFIQNDQCEMDTTDKWKLTDNGENSHSVMLKSNILYWRITGLMGSKVPV
LVKNITIEGVATSECFCKPGTFSNKPGEFNCQVPRNTYSEKAKECIRCKDSQF

SEGSSECTERPPCTTKDYFIQIHTPCDEKQTOIMYKWIPEKICREDLTDAIRUPPSG
EKDCPCNFNGFYANGSSCHPCPPGTFSDGTEKRCPCPAGTEPALGFYKWNVLPG
NMKTSFCNVGNSKCDGMNGWEVADGHQSGAGSDNDYLLNLHLPKPKPTSMGTGAT
GSELGRTEFVPELTCSADCVLYFMVDINRKNSTNVESGGTKKQAVTHIIFKNATFT
FTWGIPELQGPRT
BASE COUNT 495 a 381 c 409 g 451 t 1 others
ORIGIN

Query Match 42.2%; Score 1499.6; DB 6; Length 1737;
Best Local Similarity 98.7%; Pred. No. 0;
Matches 1545; Conservative 0; Mismatches 10; Indels 10; Gaps 3;
Qy 189 GAAAGATTATCACTTTGAATATACGAATGTGATAGCAGTGGCTCCAGGTGGAGAGTTGC 248
Db 142 GAAAGATTATCACTTTGAATATACGAATGTGATAGCAGTGGCTCCAGGTGGAGAGTTGC 201
Qy 249 CATTCCAAATTCGACGTGGACTGCTGGCTGCTGACCCAGTCGAGAGCAAGAATG 308
Db 202 CATTCCAAATTCGACGTGGACTGCTGGCTGCTGACCCAGTCGAGAGCAAGAATG 261
Qy 309 CACTTCTCTGCTTCTGGAGAGTATCTAGAAATGAAGAACCAGGTATGCAGTAAGTG 368
Db 262 CACTTCTCTGCTTCTGGAGAGTATCTAGAAATGAAGAACCAGGTATGCAGTAAGTG 321
Qy 369 TGGTGAAGCACCTATTCCTTGGGAGTGGCATCAAAATTTGATGAATGGATGAATTCGC 428
Db 322 TGGTGAAGCACCTATTCCTTGGGAGTGGCATCAAAATTTGATGAATGGATGAATTCGC 381
Qy 429 GCAGGATTTCTACATCGCAACATTCATGACACTGTGGTGGCCCTTCGACAGCAG 488
Db 382 GGCAGGATTTCTACATCGCAACATTCATGACACTGTGGTGGCCCTTCGACAGCAG 441
Qy 489 GGCAGGCTGTGAACACTCTTCTTGGATCCCTCGTGGAACTACATAGAATCTAATCG 548
Db 442 GGCAGGCTGTGAACACTCTTCTTGGATCCCTCGTGGAACTACATAGAATCTAATCG 501
Qy 549 TGATGACTGACGGTGTCTTGTATCTATGCTGTGCACCTTTAGAAGTCAGGCTATGCTTT 608
Db 502 TGATGACTGACGGTGTCTTGTATCTATGCTGTGCACCTTTAGAAGTCAGGCTATGCTTT 561
Qy 609 CTTTGTAGTACCAGTATGTCGCAACAACATCTCTTTGTAGTCTTTTATTCAAAATGATCA 668
Db 562 CTTTGTAGTACCAGTATGTCGCAACAACATCTCTTTGTAGTCTTTTATTCAAAATGATCA 621
Qy 669 GTGCCAGGAGATGGACACCACTGACAACTGACAACTGGGTAACTTACAGCAATGGAGAATG 728
Db 622 GTGCCAGGAGATGGACACCACTGACAACTGACAACTGGGTAACTTACAGCAATGGAGAATG 681
Qy 729 GGCTCTCATCTGTAATGCTGAATCAGGCACAAACATCTCTACTGGAGAACTACAGG 788
Db 682 GGCTCTCATCTGTAATGCTGAATCAGGCACAAACATCTCTACTGGAGAACTACAGG 741
Qy 789 CATCTTTATGGTCTTAAGCGGTCAAGCTGTGCTGTTGAGTCTTTTATTCAAAATGATCA 848
Db 742 CATCTTTATGGTCTTAAGCGGTCAAGCTGTGCTGTTGAGTCTTTTATTCAAAATGATCA 801
Qy 849 GTGGCGGTACACATCAAGATGTTTTCTTGCAGCCAGGCACATTCAGCAACAAACAGG 908
Db 802 GTGGCGGTACACATCAAGATGTTTTCTTGCAGCCAGGCACATTCAGCAACAAACAGG 861
Qy 909 TTCATTTCAAGTGCAGGTGTCCGAGAACACCTATCTTGACAAAGGAGCCAAAGATG 968
Db 862 TTCATTTCAAGTGCAGGTGTCCGAGAACACCTATCTTGACAAAGGAGCCAAAGATG 921
Qy 969 TATAAGGTGTAAAGAGACTCTCAATTTTC- - - - - AGGATCCAGTCAAGTGTACAGAGCG 1022
Db 922 TATAAGGTGTAAAGAGACTCTCAATTTTCAGAGGAGGATCCAGTCAGTGTACAGAGCG 981
Qy 1023 CCCTCCCTGTACCAAAAGACTATTTTCCAGATCCATCTCCATGTGATGAAGAAGAAA 1082
Db 982 CCCTCCCTGTACCAAAAGACTATTTTCCAGATCCATCTCCATGTGATGAAGAAGAAA 1041
Qy 1083 GACACAGATAATGTACAGTGGATAGAGCCCAAAATCTGCCGGGAGGATCTCAGATGC 1142

Db 1042 GACACAGATAATGTACAAAGTGGATAGAGCCCAAAATCTGCCGGGAGGATCTCAGACATGC 1101
Qy 1143 TATTAGATTGCCCTTCTTGAGAGAGAAGGATTTGTCGCCCTTGCACACCTTGGATTTTA 1202
Db 1102 TATTAGATTGCCCTTCTTGAGAGAGAAGGATTTGTCGCCCTTGCACACCTTGGATTTTA 1161
Qy 1203 TAACAATGGATCATCTTCTTGGCATCCCTGTGCTCCTCGAAGCATTTTCAGATGGAAACAA 1262
Db 1162 TAACAATGGATCATCTTCTTGGCATCCCTGTGCTCCTCGAAGCATTTTCAGATGGAAACAA 1221
Qy 1263 AGAATGTAGACCATCTCCAGAGGAACGAGCCTGCACCTTGGCTTTGAATATAAATGGTG 1322
Db 1222 AGAATGTAGACCATCTCCAGAGGAACGAGCCTGCACCTTGGCTTTGAATATAAATGGTG 1281
Qy 1323 GAATGTCTCTTCTGCGCAACATGAAACATTCCTGCTTCAATGTGGGAAATTCAAAGTGCGA 1382
Db 1282 GAATGTCTCTTCTGCGCAACATGAAACATTCCTGCTTCAATGTGGGAAATTCAAAGTGCGA 1341
Qy 1383 TGAATGAATGTTGGAGGTGGTGGAGATCATATCCAGAGTGGGGCTGGAGTTCCTGA 1442
Db 1342 TGAATGAATGTTGGAGGTGGTGGAGATCATATCCAGAGTGGGGCTGGAGTTCCTGA 1401
Qy 1443 CAATGATTACCTGATCTTAAACTTGCATATCCAGGATTTAAACCCACCAACATCTATGAC 1502
Db 1402 CAATGATTACCTGATCTTAAACTTGCATATCCAGGATTTAAACCCACCAACATCTATGAC 1461
Qy 1503 TGGAGCCAGGGTCTTGAACACTAGGAAGATAACATTTGTCTTTTGAGACCTCTGTTTCAGC 1562
Db 1462 TGGAGCCAGGGTCTTGAACACTAGGAAGATAACATTTGTCTTTTGAGACCTCTGTTTCAGC 1521
Qy 1563 TGACTGTGTTTTGTACTTTCATGTTGATATTATAGAAAAGTACAAAATGTGTAGAATC 1622
Db 1522 TGACTGTGTTTTGTACTTTCATGTTGATATTATAGAAAAGTACAAAATGTGTAGAATC 1581
Qy 1623 GTGGGTGGAACCAAGAAAACAAGCTTTACACCCATATCATCTTCAAGAAATGCAACTTT 1682
Db 1582 GTGGGTGGAACCAAGAAAACAAGCTTTACACCCATATCATCTTCAAGAAATGCAACTTT 1641
Qy 1683 TACATTTACATGGG--CATTCAGAGAACTAAT--CAGGTCAGATATAATAGACGGTTCA 1738
Db 1642 TACATTTACATGGGCAATTTCCAGAGAACTAATTTACGGGTCCAGAGATAATAGACGGTTCC 1701
Qy 1739 TCNAAT 1743
Db 1702 NCCAT 1706

RESULT 3

AX060301

LOCUS

DEFINITION

AX060301

ACCESSION

AX060301.1

VERSION

KEYWORDS

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

Shinkets,R.A., Fernandes,E., Vernel,C., Yang,M., Boldog,F.L. and

Herrmann,J.L.

TITLE

JOURNAL

FEATURES

Source

1. .1508

/organism="Homo sapiens"

/db_xref="taxon:9606"

226. .1461

/note="unnamed protein product"

/codon_start=1

/protein_id="CAC24870.1"

AX060301 Sequence 1508 bp DNA Linear PAT 22-JAN-2001

AX060301 Sequence 9 from Patent WO078802.

AX060301

AX060301.1

GI:12405790

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

Shinkets,R.A., Fernandes,E., Vernel,C., Yang,M., Boldog,F.L. and

Herrmann,J.L.

TITLE

JOURNAL

FEATURES

Source

1. .1508

/organism="Homo sapiens"

/db_xref="taxon:9606"

226. .1461

/note="unnamed protein product"

/codon_start=1

/protein_id="CAC24870.1"

/db_xref="GI:12405791"
/translation="MHSSWIPRGNVIESNRDDCTVSLIYAVHLKSGVVEFEYOYVDN
NIEFEFFIONDOCEMDTDDKWKLTONGEMSHVLMKSGTSLIYWRITGILMSK
AVKPLVKNITIEGVATISECPKPGTFSNKPQSFNCQVPRNTYSEKGAKECIRCK
DOSFSESGSECTERPCCTTKDQFQIHTPCDEEGKQIMYKWIPEKPKICREDLDIAIR
LPPSEKDKDPCPCPNFYNNSSCHPCPGTSDGTEKRCPCPAGTAPALGFYKWW
MFLPNMKTSFNVNKSQDGMNWEVAGDIQSGAGGSDNDYLILNLHIIPGPKPPTS
MTGATGSELGRITFVFEVFLCSADCVLVPMVDINRKSNTNVVESWGTKEKQAYTHIFK
NATFTFWGIPRELIOGPR"
BASE COUNT 444 a 306 c 355 g 402 t 1 others
ORIGIN

Query Match 33.3%; Score 1182.6; DB 6; Length 1508;
Best Local Similarity 98.4%; Pred. No. 7.5e-267;
Matches 1228; Conservative 0; Mismatches 10; Indels 10; Gaps 3;
QY 506 ACTCTTCTTGGATCCCTCGTGGAACTACATAGATCTAATCGTGATGACTGCACGGTGT 565
DB 230 ACTCTTCTTGGATCCCTCGTGGAACTACATAGATCTAATCGTGATGACTGCACGGTGT 289
QY 566 CTTTGATCTATCTGTGCACCTTAAGAAGTCAGGCTATGCTTCTTTGAGTACCAGTATG 625
DB 290 CTTTGATCTATCTGTGCACCTTAAGAAGTCAGGCTATGCTTCTTTGAGTACCAGTATG 349
QY 626 TCGACAACAACATCTCTTTGAGTCTTTTATTCAAAATGATCAGTCCCGAGGATGGACA 685
DB 350 TCGACAACAACATCTCTTTGAGTCTTTTATTCAAAATGATCAGTCCCGAGGATGGACA 409
QY 686 CCACCACTGACAAGTGGTAAACCTTACAGACAATGGAGAATGGGGCTCTCATTTCTGTAA 745
DB 410 CCACCACTGACAAGTGGTAAACCTTACAGACAATGGAGAATGGGGCTCTCATTTCTGTAA 469
QY 746 TCCTGAATCAGGCACAAACATACCTACTGTGAGAACCTACAGGATCCTCTTATGGTCTTA 805
DB 470 TCCTGAATCAGGCACAAACATACCTACTGTGAGAACCTACAGGATCCTCTTATGGTCTTA 529
QY 806 AGCGGCTCAAGCCTGTCTGTTGAAAAATATACAAATGAAGGGTGGCGTACACATCAG 865
DB 530 AGCGGCTCAAGCCTGTCTGTTGAAAAATATACAAATGAAGGGTGGCGTACACATCAG 589
QY 866 ATGTGTTTCTTGCAGCCAGGCACATTCAGCAACAACACAGGTTCTATCACTGCCAGG 925
DB 590 ATGTGTTTCTTGCAGCCAGGCACATTCAGCAACAACACAGGTTCTATCACTGCCAGG 649
QY 926 TGTGTCCAGAAACACCTATTCTTGAGAAAGGAGGCAAGAAATGTATAGGTGTAAGAGC 985
DB 650 TGTGTCCAGAAACACCTATTCTTGAGAAAGGAGGCAAGAAATGTATAGGTGTAAGAGC 709
QY 986 ACTCTCAATTTTC-----AGATCCAGTGTGTACAGAGCGCCCTCCCTGTACCAAA 1039
DB 710 ACTCTCAATTTTCAGAGGAAGGATCCAGTGTGTACAGAGCGCCCTCCCTGTACCAAA 769
QY 1040 AAGACTATTTCAGATCCATCTCCATCTGATGAGAGGAAGAAAGACAGATAATGTACA 1099
DB 770 AAGACTATTTCAGATCCATCTCCATCTGATGAGAGGAAGAAAGACAGATAATGTACA 829
QY 1100 AGTGATAGAGCCCAAAATTCGCGGGAGGATCTCACAGATGCTATTAGATTGCCCTTT 1159
DB 830 AGTGATAGAGCCCAAAATTCGCGGGAGGATCTCACAGATGCTATTAGATTGCCCTTT 889
QY 1160 CTGGAGAGAAGAGGATTGTCCGCCCTTGAACCCCTGGATTATTAACAATGGATCATCTT 1219
DB 890 CTGGAGAGAAGAGGATTGTCCGCCCTTGAACCCCTGGATTATTAACAATGGATCATCTT 949
QY 1220 CTTCGCATCTCTCTCTGGAACATTTTCAGATGGGAACCAAGAATGTAGACCATGTC 1279
DB 950 CTTCGCATCTCTCTCTGGAACATTTTCAGATGGGAACCAAGAATGTAGACCATGTC 1009
QY 1280 CAGCAGGAACGAGGCTGCATCTTGCTTTGAATATAAATGTGGAAATGTCTTCTCTGGCA 1339
DB 1010 CAGCAGGAACGAGGCTGCATCTTGCTTTGAATATAAATGTGGAAATGTCTTCTCTGGCA 1069
QY 1340 ACATGAAAACTTCTCTGCTTCAATGTGTGGAAATTTCAAGTGGCATGGAATGAATGTTGGG 1399

DB 1070 ACATGAAAACTTCTCTGCTTCAATGTTGGAAATCAAGTGGCATGGAATGAATGGTTGGG 1129
QY 1400 AGGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTGACAATGATTACCTGATCT 1459
DB 1130 AGGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTGACAATGATTACCTGATCT 1189
QY 1460 TAAACTTGTATATCCAGAGATTAAACCAACCAACATCTATGACTGGAGCCACGGTTCTG 1519
DB 1190 TAAACTTGTATATCCAGAGATTAAACCAACCAACATCTATGACTGGAGCCACGGTTCTG 1249
QY 1520 AACTAGAGAATAACATTTGCTTTTGGAGACCTCTGTTTCAGCTGACTGTGTTTCTACT 1579
DB 1250 AACTAGAGAATAACATTTGCTTTTGGAGACCTCTGTTTCAGCTGACTGTGTTTCTACT 1309
QY 1580 TCATGTTGATATTATAGAAAAAGTACAAATGTGTAGAACTCGTGGGTGGAACCAAG 1639
DB 1310 TCATGTTGATATTATAGAAAAAGTACAAATGTGTAGAACTCGTGGGTGGAACCAAG 1369
QY 1640 AAAACAAGCTTACACCCATATCATCTTCAAGNATCAACTTTTACATTTTACATGGG--C 1697
DB 1370 AAAACAAGCTTACACCCATATCATCTTCAAGNATCAACTTTTACATTTTACATGGGCA 1429
QY 1698 ATTCCAGAGAACTAAT--CAGGCTCAAGATAATAGACGTTTCATCAAT 1743
DB 1430 TTCCAGAGAACTAATTTACAGGTCCCAAGATAATAGACGTTTCNCCAT 1477

RESULT 4

LOCUS AX036667 3280 bp DNA linear PAT 16-NOV-2000
DEFINITION Sequence 1 from Patent WO0058460.
ACCESSION AX036667
VERSION AX036667.1 GI:11226242
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3280)
AUTHORS Bruck,C.E., Vinals,Y.D., Coche,T. and Cassart,J.P.
TITLE Novel compounds
JOURNAL Patent: WO 0058460-A 1 05-OCT-2000;
BRUCK CLAUDELINE ELVIRE MARIE (BE); SMITHKLINE BEECHAM BIOLOG (BE);
VINALS Y DE BASSOLS CARLOTA (BE); COCHE THIERRY (BE); CASSART
JEAN POL (BE)
FEATURES
source Location/Qualifiers
1..3280
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 810 a 930 c 797 g 743 t
ORIGIN

Query Match 27.9%; Score 991.4; DB 6; Length 3280;
Best Local Similarity 60.6%; Pred. No. 7.5e-222;
Matches 1758; Conservative 0; Mismatches 1116; Indels 27; Gaps 7;

QY 171 GCTTCTCTTCCAGGAGAAAGATTATCATTGTAATATACGGAATGTGATAGCAGTGG 230
DB 135 GCTTCTCTTCCAGGAGAGTCTGAGTACCTATGAGTACACGGGCTGTGACACGACGG 194
QY 231 CTCCAGGTGGAGAGTTGCCATTTCCTGAGTGGACTGCTCTGGGCTCGCTGACCC 290
DB 195 TTCCAGGTGGAGGGTCCGCGTCCGACCTACCCGGGCGCTGTGCACGAGCTGCCCTGACCC 254
QY 291 AGTGAGAGCAAGAATGCACCTTCTCTCTGCTGCTTCTGGAGAGTATCTAGAATGAAGAA 350
DB 255 CGTGAAGGCAAGGAGTCTCTCTCTCTGCAACCGCGGGAGTCTCTCGATATGAAGA 314
QY 351 CGAGTATGAGTATGAGTGTGGAAGGACCTATTCTTGGGAGTGGGACCAATTTGA 410
DB 315 CGAGTATGAGTATGAGTGTGGAAGGACCTATTCTCTCGGCGGCTACTCTCTCGGACAGGCAATTCGGTTTCA 374

Qy 411 TGAATGGATGAATGGCCGACAGGATTTTCTAACATCGCAACATTCATGGACACTGTGGT 470
Db 375 TGAGTGGATGAGCTGCCCATGGCTTGGCCAGCTTCAGCCCAACATGGAGCTGGATGA 434
Qy 471 GGGCCCTTCTCAGACAGCCGACAGCGCTGTAAACACTCTCTTTGGGATCCCTCGTGGAA 530
Db 435 CAGTGTCTGAGTCCA---CCGGGAAGTGTACTTCGTCCTCAAGTGGGTCCCGGGGGGA 491
Qy 531 CTACATGAATTAATCGTGATGACTGCACGGTGTCTTTGATCTATGCTGTGCACCTTAA 590
Db 492 CTACATGCCCTCCACACGGGACGATGCACAGCCACACTGATGATACGCCGTCAACCTGAA 551
Qy 591 GAAGTCAGGCTATGCTCTTTTGGTACGAGTATGTCGACAAACATCTCTTTTGGAGTT 650
Db 552 GCAATCTGGCCGTTAACTTCGAATACTACTATCCAGACTCCAGCATCTTTTGGATT 611
Qy 651 CTTTATTCAAAATGATCAGTGGCAGGAGATGGACACCACCACTGACAAAGTGGTAAACT 710
Db 612 TTTCCTTCAGAATGACCACTGCCAGCCCAATGCAGATGACTC---CAGGTGGATGAAGAC 668
Qy 711 TACAGACAATGGAGAAATGGGGCTCTCAATTCCTGTAATGCTGAATCAGGCACAAACATACT 770
Db 669 CACAGAGAA---AGATGGGAATCCACAGTGTGGAGCTAAATCGAGGCAATATGTCTT 725
Qy 771 CTACTGGAGAACTACAGGCATCTTTATGGTTCCTAAGCGGTCAAGCCGTGCTGGTAAA 830
Db 726 CTATTGGAGAACCAAGGCTCTCAGTATGGACCAAAAGTACCACCAAGCCGTGCTGGTGA 785
Qy 831 AAATATCACAATTAAGGGTGGCTTACATCACAATCTTTTCTCTCAAGCCAGGCAC 890
Db 786 AAACATTCGCAATTAAGGGTGGCTTACATCTCAAAATGCTTCCCTCGCAACCTTGGCAC 845
Qy 891 ATTACAGCAACACAGGTTTCAATCAACTGCCAGGTGTGTCCAGAAACACACTATTCTGA 950
Db 846 GTATGCAGACAGCAGGGCTCTCTTTCTGCAAACTTTGCCCGAGCAACTTTATTCAA 905
Qy 951 GAAAGAGCAAGAAATGATTAAGTGTAA---AGACGACTCTCAATTTTCAGGATCCAG 1007
Db 906 TAAAGGAGAAACTTTCTGCCACCACTGTGACCTTGACAAATACTCAGAGAAAGGATCTTC 965
Qy 1008 TGAGTTCAGAGAGCCCTCCCTGTACCAACAAAGACTATTTCCAGATCCATCTCCATG 1067
Db 966 TTCCCTGTACGTGGCCCGAGCTTGCACAGACAAGATTAATTTCTACACACACACGCCCTG 1025
Qy 1068 TGATGAAGAAGAACACACAGATAATGTACAAGTGGATAGAGCCCAAAATCTGCGGGA 1127
Db 1026 CGATGCCAACGGAGAGACACAACATCATGTACAAATGGGCCAAGCCGAAATCTGTAGCA 1085
Qy 1128 GGATCTCAGATGCTATTAGATTGCCCCCTTCTGGAGAGAGAGGATTTGCCGCTTG 1187
Db 1086 GGACCTTGAGGGGGCAGTGAAGCTGCTGCTGCTGTGTAAGACCCACTGCCACCCCTG 1145
Qy 1188 CAACCTGGATTTTATAACAATGGATCATCTTCTGCCATCCCTGCTCTCTGGAAACAT 1247
Db 1146 CAACCCAGGCTCTTCCAAACCAACACAGACCTGCCAGCCCTGCCATATGTTCCCTA 1205
Qy 1248 TTCAATGAACCAAGAAATAGACCATGTCCAGCAGGAAGCGGCTGCACTTGGCTT 1307
Db 1206 CTCCAATGG---CTCAGACTGTACCCGCTGCCCTGCGAGGACTGAACTGCTGTGGGAT 1262
Qy 1308 TGAATATAAATGGTGAATGTCTCTCTGCGACATGAANAATCTCTGCTTCAATGTTGG 1367
Db 1263 TGAATACAATGGTGAACACGCTGCCCAACAACATGGAAAGCAGCGTCTCTCAGTGGGAT 1322
Qy 1368 GAATTCAAAGTGGGATGAATGAATGGTGGAGGTGGCTGGAGATCATATCCAGAGTGG 1427
Db 1323 CAATTCGAGTACAGGGATGACAGGCTGGGAGGTGGCTGGTGGATGATCATATTTACACAG 1382
Qy 1428 GGCCTGGAGTTCAGACATGATTTACTGATCTTAACTTGAATATCCAGGATTTAAACC 1487
Db 1383 TGCTGGAGCCTCAGACAAATGACTTTCATGATTTCTCACTCTGTTGTGCCAGGATTTAGACC 1442
Qy 1488 ACCAACATCTATGACTGGAGCCAC---GGGTTCTGAACTAGGAAGAAATACATTTGCTT 1544

Db 1443 TCCGAGTGGTGTATGGCAGACACAGAGAATAAAGAGGTGCCAGATTCATTTTGTCTT 1502
Qy 1545 TGAGACCCCTCTGTTCAGCTGACTGTGTTTGTACTTTCATGTGGATATTAATAGAAAAAG 1604
Db 1503 TGAGACCCCTCTGTTCGTGAACCTGTGAGCTCTACTTTCATGTGGTGTGAATTTCTAGGAC 1562
Qy 1605 TACAAATGTGTAGANTCGTGGGTGGAAACCAAGAAACCAAGCTTACACCCATATCAT 1664
Db 1563 CAACACTCTCTGGAGAGCTGGAAGGTTCCAAAAGGCAACAGCTCTATACCTACATCAT 1622
Qy 1665 CTTTCAAGAAATCGAACTTTTACATTTTACATGGGCAATTCAGAGAACTAATCAGGTTCAAGA 1724
Db 1623 TGAGGAGAACTACTACAGAGCTTCCACCTGGCCCTTCAGAGGAGCCACTTTTTCATGAGGC 1682
Qy 1725 TAATAGACGGTTCATCAATGACATGGTGAAGATTTTATTTCTATCAGAGCCACTTAATGCAAT 1784
Db 1683 AAGCAGGAAGTACACCAATGAGTTTGGCAAGATCTACTCCATCAATGTCCACCAATGTTAT 1742
Qy 1785 TGATGGGTGGGCTCTCATGCGGTGCCCTGTGCCCTCGGTTCTGAACAGTCCGGTTTCATC 1844
Db 1743 GAATGGCTGGGCTCTCTACTGCGGTCCCTGTGCCCTTAGAAGCCTCTGATGTGGGCTCTCTC 1802
Qy 1845 CTGTGTCCCTGCCCTCCAGGCCACTTACATTCAGAAAGAAACCAACCACTGCAAGGAATG 1904
Db 1803 CTGCACCTCTTGTCTGCTGCTGTTACTATATTGACCGAGATTCAGGAACCTGCCACTCCTG 1862
Qy 1905 TCCACCTGACACCTACCTGTCCATACATCAGTCTATATGGCAAGAGGCTGTGATTTCCATG 1964
Db 1863 CCCCCCTTAACACAATTCCTGAAGAGCCACAGCCCTTATGGTGTCCAGGCTGTGTGCCCTG 1922
Qy 1965 CGGGCTCGGAGTAAACAACTCAGACCAATTCGGTTTGTCTATAGTACTGCTTTTCTTA 2024
Db 1923 TGGTCCAGGGACCAAGAACCAAGATCCACTCTCTGTGCTACAATGATGTGACCTTTCTC 1982
Qy 2025 CCATGAAAAGAAATCAGATTTTGCACCTATCCTATCTAGCAACCTCAGCAGTGTGGGCTC 2084
Db 1983 AGCAACACTCCACAGGACTTTCAACTACAACTTCTCCGCTTTGGCAACACACCTCAC 2042
Qy 2085 ATTAATGAATGGCCCGCCAGCTTCACCTCCAAAGGAACAATACTCTCCATTTCTTCAATAT 2144
Db 2043 TCTTGTGTGGAGGCAAGCTTCACTTCCAAAGGTTTGAATACTTCCATCCTTTACCGCT 2102
Qy 2145 CAGTTTATGTGGCATGAGGGGAAGAGTGGCTCTCTGTACCAACAATAATACAGACTT 2204
Db 2103 CAGTCTCTGTGAAACACAGGTTAGGAAATGCTGTGTGCACCCGAGCAATGTCTACGTACCT 2162
Qy 2205 TACAGTAAAGAAATAGTGGCAGGTCAGATGATACACAAATTTTGGTAGGGCATTTTGT 2264
Db 2163 CCGGATTCCTGAGGGTGAAGTCAAGG-----TTCCTCCAAATCTATCACAGCCTACGT 2213
Qy 2265 ATGCCAGTCAACAATTAATTTCTTCTGAAAGTAAAGGTTTCCGAGCAGCCTTATCATCACA 2324
Db 2214 CTGCCAGGCAGTCACTATCCCCCAGAGGTCAGAGGCTACAAGCGGGGTTTCTCTCACA 2273
Qy 2325 ATCCATCATCTTGGCAGATACATTCATAGAGTCAAGTGAACCAACATATGAAAAATAT 2384
Db 2274 GCTGTGTCAGCCTTGTGTCGAGCTTATTTGGGGTGAACACAGATATGACTCTGGATGGAAT 2333
Qy 2385 TAATATAAAGAAAGATATGTTCCACGTTCCAAACAGCCAAATACCAAGATGTGATTTCTT 2444
Db 2334 CACCTCCCGAGCTGAATTTTCCACCTGGAGTCCCTTTGGGAATACCGGAGCTGATCTTCTT 2393
Qy 2445 TTATAAGTCTTACAGCAACAACATCTTGTATTAATGGCCGATCAACTGCTGTGAAAT 2504
Db 2394 TTATAGTCCAATGATCTGACCCAGTCTTCCAGTTCTGGGAGATCAACCCACCTCCCGCT 2453
Qy 2505 GAGGTCTAATCTTACTAAATCTCGAGCAGGAGTATTTTCAAGTCCCGCAGCAAGTGGCCAGC 2564
Db 2454 CAGGTGAGTCCACAGAAACTCTCCCTGGAAGTTTGTGCTGCCAGGACGTGCTCAGA 2513
Qy 2565 AGGTAGCTGTGATGGGTGTACGTTCTATTTCTCTGTGGGAGAGTGTGAAAGCTTGGCCCTCT 2624

Db 2514 TGGGACCTGTGATGGGTGCAACTTCCACTTCTGTGGGAGCGCGCTGCTTGCCTGCT 2573
QY 2625 GTGTAGGAGCATGACTTCCATGAGATTTAGGGAGCTCTCAAGAGAGGATTTTCAGGAAC 2684
Db 2574 CTGCTCAGTGGCTGACTACCATGCTATGCTCAGCAGCTGTGTGGCTGGGATCCAGAGAC 2633
QY 2685 CTTGTATGTGGAATGAACCTAAATGGTGCAATTAAGGAATTTCTTTGCCCTGAGAAAA 2744
Db 2634 TACTTACGTGTGGCGAAGACCAAGCTATGCTCTGGTGGCATTTCTCTGCCTGAGCAGAG 2693
QY 2745 GTTGGCAACCTGTGAACGGTTGACTTTTGGCTGAGGTGGGAGCGGTGTGGGACCTTT 2804
Db 2694 AGTCACCATCTGCAAAACCATAGATTTTGGCTGAAAGTGGGCATCTCTGAGGACCTTG 2753
QY 2805 TACTGCGCTTTGCTGGTGGCTGCTGACCTGCTACTTCTGGAAAAAACAATAAACTGGA 2864
Db 2754 TACTGCCATCTGCTCACCCTCTTGACCTGCTACTTTTGGAAAAAAGAACTAAGTAGA 2813
QY 2865 ATACAATATTCCAAGTTAGTATGACGACTTAACCTCAAGAGAGTGAACCTCCCGGCTGC 2924
Db 2814 GTACAAGTACTCCAAGCTGGTGATGAATGCTACTCTCAAGGAGCTGTGACCTGCCAGCAG 2873
QY 2925 AGACAGTTGTGCTATCATGGAAGGAGAAATGAATGAAGAGGAGTTGTATATTCCATTA 2984
Db 2874 TGACACTGCGGCATCATGGAAGCGAGAGTGTAGAGGAGCAGCTCATCTTTACCGACAA 2933
QY 2985 ACAGTCACTACTAGGAAAACTCAAAATCTTTGGCAACCAAGAAAAAGAACCACTTTTGA 3044
Db 2934 GAAGTCACTCTTTGGGAAGTCAAAATCAITTACCTCCAAGAGAGCTCTCTGGGATTGA 2993
QY 3045 ATCTGTTCAACTGMAAACCTC 3065
Db 2994 CTCAGTGGCGCTGAGAGACATC 3014

RESULT 5
AX12726
LOCUS AX12726 3331 bp DNA linear PAT 15-MAY-2001
DEFINITION Sequence 1 from Patent WO0131003.
ACCESSION AX12726
VERSION AX12726.1 GI:14134373
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 3331)
JOURNAL Delneste,Y., Magistrelli,G., Jeannin,P. and Bonnefoy,J.Y.
Cloning, expression and characterisation of a gene expressed in
tumour cells and involved in the regulation of the immune response
Patent: WO 0131003-A 1 03-MAY-2001;
PIERRE FABRE MEDICAMENT (FR)
FEATURES
Location/Qualifiers
1..3331
/organism="Homo sapiens"
/db_xref="taxon:9606"
45..3086
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC38907.1"
/db_xref="GI:14134373"

/translation="MAEPGHSHLSARVGRTERIPRLWRLWAGTAFQVGTGTP
ELHACKSEYHYEYACDSTGRWRVAPHTPLGCTSLDPVKGTGECFSNAGEFLD
MKDSCRPCAEGRYSLGTGIRFDEWDELPHGASLANNELDDSAESTGTSKXW
PRGDIYASNTDECTATLMYAVNLKQSGTVNFYYPDSIIIFEFVNDQCPNADDS
RWMTKEGWFEHSEVNLGRNNVLYWRITAFSVMTKPKPVLVRLNIATIGVAYSECF
PKPCTYADKQSGSFCMLKCPANSYNSKGETSCHQCDPKYSEKSSCNVRPACTDKD
YFYTHACDANGESTQLMKWAKPKICSEDLGAVKLPASGVKTHCPNCPGFFETNNS
TCQCPYGPYNSGDSCTPCPAGTEPAVGFYKWNLTPLNMTETVLSGINFEYKGMTG
WEVAGDHIYTAAGSDNDFMLITLVPGFRPPQSMADTENKVARITFVETLCSVN
CELYFMVNSRTNPTVETWKGSKQSYVLIENITTTSTFAFORTTFHEASPKYT
NDVAKIYSINTVNMVNGVASCRCALASDVSGSCTSCPAGYIDRDSGCHSCPNN
TILKAHQPIGVQACVPCPGPGTKNNKIHSCLYNDCTFSRNTPTRTFNFSALANTVTL

AGGPTSFKGLAVFHHFTLLSCNGRKMVCTDNVTDLRIPEGESGFSKITAYVCO
AVIIEPVTGYKAGVSSOPVSLADRLIGVTTDMTLDGITSPAELFHLESLGIPDVIFP
YRSDVTQSCSSGRSTTIRVRSPOKTYFPGSLPGLTSGDGTGCGDNHFLWESAAC
PUCSVADYRAIVSSCVAGIKRTYVWREPKLCSGGISLPEQRVTICKTIDFLMKVGIS
AGCTAAILTLVTLCYFWKKNKQLEYKSKLVNMNATLKDCLDPAADSCAIMEGEDVEDD
LIFTSKKSLFGIKSKFTSKRTDPGDFSVPLKTSSTGGPMDL"

BASE COUNT 822 a 945 c 813 g 751 t
ORIGIN
Query Match 27.8%; Score 989.8; DB 6; Length 3331;
Best Local Similarity 60.6%; Pred. No. 1.8e-221;
Matches 1757; Conservative 0; Mismatches 1117; Indels 27; Gaps 7;

QY 171 GCTTCTCTCTTGCAGGAGAAAGATTATCACTTTGAATATACGAATGTGATAGCAGTGG 230
Db 179 GCTTCAATGCTGCAAGAGTCTGAGTACCACTATGAGTACACGGCGTGTGACAGCAGGG 238
QY 231 CTCAGGTGGAGAGTTGCCATTCCAAATTCGCAAGTGGAGTCTGTGGCCTGGCTGACCC 290
Db 239 TCCAGGTGGAGGTGCGCGTCCGCATACCCCGGGCTGTGCACCAAGCTGCTGACCC 298
QY 291 AGTGAGAGCAAGAATGACATTTCTCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAA 350
Db 299 CGTCAAGGGCACCGAGTGTCTTCTCTGCAACGCCGGGAGTTCTTGATATGAAGGA 358
QY 351 CAGGTATGCAAGTGTGTCAGGCACTATTCTCTGGGAGTGGCATCAAAATTTGA 410
Db 359 CCAATGATGTAAGCCATGCGCTGAGGGCGCTACTCCTCGGCACAGGATTCGGTTGA 418
QY 411 TGAATGGGATGAATTCGGCGAGGATTTTCAACATCGCAACATTCATGACACATCTGGT 470
Db 419 TGAGTGGGATGAGTGCCTCATGGCTTTGCCAGCCTCTCAGCAACATGAGCTGGATGA 478
QY 471 GGGCCCTTCTGACAGCAGGCGGCTGTGAACACTCTTCTTGGATCCCTCGTGAAA 530
Db 479 CAGTGTGCTGAGTCA---CCGGGAACCTGTACTTGTCCCAAGTGGTTCCCGGGGCGA 535
QY 531 CTACATAGAACTTAATCTGTGACTGACGCGTGTCTTTGATCTATGCTGTGACCTTAA 590
Db 536 CTACATCGCTCCCAACACGAGCAATGCAGCCACACTGATGACCGCTCAACCTGAA 595
QY 591 GAAGTCAGGCTATGCTTCTTTGAGTACCAGTATGTGCAACAACATCTTCTTTGAGTT 650
Db 596 GCAATCTGCGACCGTTAACTTCGAATACTACTATCCAGACTCCAGCATCATCTTTGAGTT 655
QY 651 CTTATTCAAAATGATCAGTGCAGGAGATGGACACCACTGCAAGTGGGTAAACT 710
Db 656 TTTCTGTTTCAAGATGACAGTGCAGGCCCAATGCAGATGACTC---CAGGTGGATGAAGAC 712
QY 711 TACAGACAATGGAGAATGGGGCTCTCATTTCTGTAATGCTGAAATCAGGCAACAACATCT 770
Db 713 CACAGAGAA---AGGATGGGAATTCACAGCTGTGGAGCTAAATCGAGGCAATATGCTCT 769
QY 771 CTACTGGAGAACTACAGGATCCTTATGGTCTTAAGGGGGTCAAGCCTGTGCTGTGATAA 830
Db 770 CTATTGGAGAAACACACAGCCTTCTCAGTATGGACCAAGTACCCAAGCCTGTGCTGTGAG 829
QY 831 AAATATCACAATTTGAAGGGGTGCGGTACACATCAGATGTTTCTTCTTGCAGCAGGCAC 890
Db 830 AAACATTTGCATTAACAGGGGTGGCTTACACTTTCAGAAATGCTTCCCTTGCAGAACTGGCAC 899
QY 891 ATTACAGCAACAACACAGGTTTCAATTCACCTGCCAGGTGTGCCAGCAAAACACTTCTCTGA 950
Db 890 GTATGACAGACAAGCAGGCTCTCTTCTTGCACAACTTTGCCAGCAACTCTTATTCAAA 949
QY 951 GAAAGAGCAAGAATGTATAGGTGATA---AGACGACTCTCAATTTTTCAGGATCCAG 1007
Db 950 TAAAGGAGAAACTTCTTGGCACCAGTGTGACCTGCAAAATACTCAGAGAAAGGATCTTC 1009
QY 1008 TGAGTCTACAGAGCGCCCTCCCTGTACCACAAAGAGTATTTTCCAGATCCATCTCATG 1067
Db 1010 TTCCTGTAAGCTGCGCCCGAGCTTGCACAGACAAAGATTATTTCTACACACACAGCGCTG 1069

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

Baker, K.P., Beresini, M., Deforge, L., Desnoyers, L., Fillaroff, E.,
Gao, W.Q., Geritsen, M.E., Goddard, A., Godowski, P.J., Gurney, A.L.,
Sherwood, S., Smith, V., Stewart, J.A., Tamas, D., Watanabe, C.K.,
Wood, W.L. and Zhang, Z.

Secreted and transmembrane polypeptides and nucleic acids encoding
same

JOURNAL

Patent: WO 0140466-A 37 07-JUN-2001;

FEATURES

source

Genentech Inc. (US)

Location/Qualifiers

1..3501

/organism="Homo sapiens"

/db_xref="taxon:9606"

BASE COUNT

921 a

979 c

823 g

776 t

2 others

7;

Indels

27; Gaps

7;

Matches

1755; Conservative

0; Mismatches

1119; Indels

27; Gaps

7;

Query Match

27.8%; Score

987.8; DB

6; Length

3501;

Best Local Similarity

60.5%; Pred.

No. 5.3e-221;

Matches

1755; Conservative

0; Mismatches

1119; Indels

27; Gaps

7;

Query Match

27.8%; Score

987.8; DB

6; Length

3501;

Best Local Similarity

60.5%; Pred.

No. 5.3e-221;

Matches

1755; Conservative

0; Mismatches

1119; Indels

27; Gaps

7;

Query Match

27.8%; Score

987.8; DB

6; Length

3501;

Best Local Similarity

60.5%; Pred.

No. 5.3e-221;

Matches

1755; Conservative

0; Mismatches

1119; Indels

27; Gaps

7;

Query Match

27.8%; Score

987.8; DB

6; Length

3501;

Best Local Similarity

60.5%; Pred.

No. 5.3e-221;

Matches

1755; Conservative

0; Mismatches

1119; Indels

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86

87

88

89

90

91

92

93

94

95

96

97

98

99

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

```
QY 2025 CCATGAAAAAGAAATACAGATTTTCCACATATGACATTTTACGAAACCTCAGCAGTGTGGGCTC 2084
Db 2115 ACGCAACACTCCACAGGAGACTTTTAACTTCTCCGCTTTGGCAACACACGCTCAC 2174
QY 2085 ATTAATGAATGCCGCCAGCTTCCACCTCCAAAGGAACAATACTTCCATTTCTTCAATAT 2144
Db 2175 TCTTGGTGGAGGCCAGCTTCCACTTCCAAAGGGTTGAATACCTTCCATCCTTACCCCT 2234
QY 2145 CAGTTTATGTGGCGATGAGGGAAGAAGTGGCTCTCTGTACCAACAATATATACAGACTT 2204
Db 2235 CAGTCTGTGGAACACAGGTTAGGAAATGCTGTGTGCACGACATGTACACTGACCT 2294
QY 2205 TACAGTAAAGAAATAGTGGCGAGGTGAGATGATATACACAATTTGGTAGGGGATTTGT 2264
Db 2295 CCGGATTCCTGAGGGTGAGTCAGGG-----TTCTCCAAATATATACAGACGCTAGCT 2345
QY 2265 ATGCCAGTCAACAATATTTCTTCTGTGAAAGTAAAGGTTTCCGACGAGCTTATATCATACA 2324
Db 2346 CTGCCAGGAGTCAATCCCCCCAGAGGTGACAGGCTACAAAGCGGGTTTCTCTACA 2405
QY 2325 ATCCATCATTTCTGGCAGATACATTTAGGAGTACAGTTGAAACACATTTGAAATAT 2384
Db 2406 GCCTGTGAGCCTTGTGATCGACTTATTGGGCTGACACAGATATGACTCTCGATGGAAT 2465
QY 2385 TAATATAAAGAGATATGTTCCAGTTCACACAGCCAAATACCAGATGTCATTTCTT 2444
Db 2466 CACCTCCCCAGTGAACCTTTTCCACTCGAGTTCCTGGGAATACCGGAGCTGATCTTCT 2525
QY 2445 TTATAAGTCTTCTACAGCAACAACATCTGTATTAATGCCCCATCAACTGCTGTGAAAT 2504
Db 2526 TTATAGTCCAATGATGACCCAGCTCTGCGAGTTCTTGGGAGATCAACACCATCCGGT 2585
QY 2505 GAGGTGTAATCTACTAATCTGGAGCAGAGTGAATTTAGTCCCCAGCAAGTGCACG 2564
Db 2586 CAGGTGCAGTCCACAGAAACTGTCCCTGGAAGTTGTGCTGCCAGGAACGTGCTCAGA 2645
QY 2565 AGTACCTGTGATGGGTGATGCTTCTTATTTCTGTTGGAGAGTGTGAGCTTGCCCTCT 2624
Db 2646 TGGGACCTGTGATGCTGCAACTTCCACTTCTGTGGGAGAGCGGCTGTGTCGCCGCT 2705
QY 2625 GTGTACGGAGCATGACTTTCATGAGATTGAGGAGCGCTGCAAGAGAGGATTTTCAGGAAC 2684
Db 2706 CTGCTCAGTGGCTGACTACCATGCTATCGTCAGCAGCTGTGTGGCTGGATCCAGANGAC 2765
QY 2685 CTGTATGTGTGGAATGAACCTTAATGTTGTCATTAAGGAATTTCTTGGCTGAGAAAAA 2744
Db 2766 TACTTACGTGTGNCAGAACCCAGCTATGCTGTGTGGCATTTCTCTGCTGTGAGCAGAG 2825
QY 2745 GTTGGCAACCTCTGAAACGGTTGACTTTTGGCTGAAGTGGGAGCGGCTGTGGAGCTTT 2804
Db 2826 AGTCACCATCTGCAAAACCATAGATTCTTGGCTGAAAGTGGGCACTCTCTGCAAGGACCTG 2885
QY 2805 TACTGCGCTTTTGGCTGTGCTGCTGACCTGCTACTTCTGAAAAAGAAATCAAAAACTTGA 2864
Db 2886 TACTGCCATCTCTGCTCAGCGTCTTGAACCTGCTACTTTTGGAAAAAGAAATCAAAAACTGA 2945
QY 2865 ATACAATAATCCAGTTAGTATGACGACTAACTCAAAAGAGTGTGAACCTCCCGGCTGC 2924
Db 2946 GTACAAGTACTCCAAGCTGGTGAATGCTACTCTCAAGGAGCTGTGACCTGCCAGCAGC 3005
QY 2925 AGCAGTTCGTCTATCATGGAAGGAGAGATTAATGAAGAGGAAGTTGTATATATCCANPAA 2984
Db 3006 TGACAGCTCGGCCATCATGGAAGGGGAGGAGTGTAGAGGACGACCTCATCTTTTACAGCAA 3065
QY 2985 ACAGTCACTACTAGGAAAACTCAAAATCTTTGGCAACCAAGGAAAAAGAACCATTTTGA 3044
Db 3066 GAAGTCACTTTTGGGAAGATCAATCATTTTACCTCCAGAGAGGACTCTCTGATGGATTGA 3125
QY 3045 ATCTGTTCAACTGAAACCTC 3065
Db 3126 CTCAGTGGCGCTGAAGACATC 3146
```

```
RESULT 7
LOCUS AX127742
DEFINITION Sequence 17 from Patent WO0131003.
ACCESSION AX127742
VERSION AX127742.1 GI:14134389
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 2610)
AUTHORS Delneste,F., Magistrelli,G., Jeannin,P. and Bonnefoy,J.Y.
TITLE Cloning, expression and characterisation of a gene expressed in
JOURNAL tumour cells and involved in the regulation of the immune response
PATENT: WO 0131003-A 17 03-MAY-2001;
PIERRE FABRE MEDICAMENT (FR)
FEATURES
Location/Qualifiers
1..2610
/organism="Homo sapiens"
/db_xref="taxon:9606"
<1..>2610
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC38913.1"
/db_xref="GI:14134390"
/translat="TGPELHACKSEYHYEYACDSTGSRWRVAVPHTPGLCTSLPDP
VKGTECSFSCNAGFELDMKQSCPCAEGRYSLGTGRFDEWDELPHGFASLSANMEL
DSSAESTGNCSTSSKWPGRDY IASNTDECTATLMYAVNLKQSGTVAFEYYPDSII
FEFTVONDQCPNADDSRWKMTTEKGHEFHSVELNRGNVLYWRTTAFSVMTKVPKV
LYRNIAITGVAYTSECPCKPTCYADKQGSFCKLCPANSYNGEPTSCUCDDPKYS
EKGSSSNVRPACTDKDYFYHTTACDANGETOLMTKWKPKICSEDLEGAVKLPASVG
KTHCPGCPNGEFTNNSTQCPYPGYSNGSDCTRCPCAGTEPAPGVFEYKWNLTPTNM
ETTVLSGTNFEYKMTGMEVAGDHITYAAGANDDMILLTLVFPGRPPQSMADTEN
KEVARIITFEETLCSVNCLELYEYVNSRTNTPEVETWKGSKQSYTVIIGENTTSP
TMAQRTTAEASRKYTNDAKIVSINVTNMGVASYCRPCALEASDVGSSCTSCPA
GYIDRSGTCHSCPPNTILKAHQPYQVQACVPCGPGTKNNKIISLCVNDCTFSRNT
TRTNYNFSALANTVTLAGSPFTSKGLKYFHFHTLSLCGNGRKMVSCTDNTVDLRI
PEGESGFSKSIYAVQAVIIPPEVTGYKAGVSQPSVLADRLGLGVDTMDLDTSP
AELFLHSLGIPDIVFFYRSDNTVQSCSSRSRTTIRVCSPOKTVPGSLLLPGTCS
TCDGNFHLWESAAACPLCSVADYRAIVSSCVAGIOKTYTVWREPKLCSGISLPEQ
RVTTCKTIDFWLVK"
BASE COUNT 634 a 744 c 644 g 588 t
ORIGIN
Query Match 24.2%; Score 860.8; DB 6; Length 2610;
Best Local Similarity 60.0%; Pred. No. 3.5e-191;
Matches 1571; Conservative 0; Mismatches 1022; Indels 27; Gaps 7;
QY 171 GCTTCTCTTCCCAGGAGAAAGATTATCACATTTCATATACGGAATGTGATAGCAGTGG 230
Db 12 GCTTCATGCTTCCAAAGAGTCTGAGTACCCTATGAGTACACGCGGTGTGACAGCACGG 71
QY 231 CTCCAGGTGGAGGTGGCGCTGCCGATACCCCGGCCCTGTGCACAGCCTGCCCTGACCC 290
Db 72 TTCCAGGTGGAGGTGGCGCTGCCGATACCCCGGCCCTGTGCACAGCCTGCCCTGACCC 131
QY 291 AGTGAGAGCAAGAATGCACATTTCTCTGTGCTCTTCTGGAGAGTATCTAGAAATCAACAA 350
Db 132 CGTCAAGGCGACCGAGTGTCTCTCTCTGCAACGCCGGGGATTTCTGGATATGAAGA 191
QY 351 CCAGGTATGCAGTAAAGTGTGGAAGGACCTATTTCCTTGGGCGAGTGGCATCAAAATTGA 410
Db 192 CCAGTCATGTAAAGCCATGGCTGAGGGCGCTACTCCCTCGCACAGGCAATTCGGTTTGA 251
QY 411 TCAATGGGATGAATTCGCGGAGAGATTTTCTAACATPCGCAACATTCATGCACACTGTGCT 470
Db 252 TGAGTGGGATGAGTGCCTCCATGGCTTTGCCAGCCTCTCAGCAACATGGAGCTGGATGA 311
QY 471 GGGCCCTTCTGACAGCAGGCCAGCGGTGTAAACAACCTCTCTTGGATCCCTCGTGGAAA 530
Db 312 CAGTGTCTGTGAGTCCA---CCGGGAACGTACTTCGTCCTCAAGTGGGTTCCTCCCGGGCGCA 368
```

Qy	531	CTACATAGAAATCTAATCGTGATGACTCGACGGTGTCTTTGATCTATGCTGTGCACCTTAA	590
Db	369	CTACATCGCTCCAAACACGGAGCAATGACAGCCACACATGATGTACGCGTCAACCTGAA	428
Qy	591	GAAGTCAGGCTATGTCTTCTTTGAGTACCGATGTGTCGACAAACAATCTTCTTTGAGTT	650
Db	429	GCAATCTGGCACCGTTAACTTCGTAATCTACTATCCAGACTCCAGCATCATCTTTGAGTT	488
Qy	651	CTTTATTTCAAAATGATCAGTGCAGGAGATGGACACCACCACCTGACAAGTGGTGAATAACT	710
Db	489	TTTCTGTCAGATGACCACTGCCAGCCCAATGCAGATGACTC---CAGGTGGATGAAGAC	545
Qy	711	TACAGACAATGGAGAAATGGGGCTCTCATCTTGTGAATGCTGAAATCAGGCACAAACATCT	770
Db	546	CACAGAGAA---AGGATGGGAATTCACAGTGTGGAGTAAATCGAGCAATATGTCCT	602
Qy	771	CTACTGGAACTACAGCATCCCTTATGGTTCCTAAGGCGGTCAAGCGCTGAGCTGTGTA	830
Db	603	CTATTGGAGAACACAGGCTTCTCAGTATGGACCACAAAGTACCCAAAGCCTGTGCTGTGAG	662
Qy	831	AAATATCACAATTTGAAGGGTGGCTACACATCAGAATGTTTCTTGCAGCGCAGGCAC	890
Db	663	AAACATGGCCATAACAGGGTGGCTACACTTCAGATGCTTCCCCTGCAAACTGGCAC	722
Qy	891	ATTACGACAAACACCAGTTCATTCACTGCCAGGTGTGTCCAGAAACACCTTATTCTGA	950
Db	723	GTATGCAGACAAGCAGGGCTCTCTTTCTGCAAACTTTGCCAGCCAACTCTTATTCAAA	782
Qy	951	GAAGAGGCAAGAATGTATAAGGTGTA---AGACGACTCTCAATTTTCAGATCCAG	1007
Db	783	TAAAGAGAAACTTCTTGCACACAGTGTGACCCCTGCACAAATCTCAGAGAAAGGATCTTC	842
Qy	1008	TGAGTTGTACAGAGCGCCCTCCCTGTACCACAAAAGACTATTTCACAGATCCATCTCCATG	1067
Db	843	TTCTGTAGCTGCGCCAGCTTGCACAGACAAGATTTATTCTACACACACAGGCGCTG	902
Qy	1068	TGATGAAGAAAGAAAGACACAGATAATGTACAAGTGGATAGAGCCCAAAATCTGCCGGGA	1127
Db	903	CGATGCCAACGGAGAGACACAACACTCATGTACAAATGGGCCAAGCCGAAATCTGTAGCGA	962
Qy	1128	GGATCTCACAGATGCTATTAGATTGCCCCCTCTGGAGAGAGAAGATTTGTCGCGCTTG	1187
Db	963	GGACCTTGAGGGGACAGTAAGCTGCCCTCTGGTGTGAAGACCCACCTGCCCCACCTG	1022
Qy	1188	CAACCCCTGGATTTTATAACAATGGATCATCTTCTTGCCATCCCTGTCTCTCTGCAACATT	1247
Db	1023	CAACCCAGGCTCTTCAAAACCAACAACGACACCTGCCAGCCCTGCCCATATGTCCTTA	1082
Qy	1248	TTCCAGATGGAAACAAAGAATGTAGACCATGTCCAGCAGAAACGAGGCTGCATTTGGCTT	1307
Db	1083	CTCCAATGG---CTCAGACTGTACCCGCTGCCCTGCAGGGACTGAACTGCTGTGGGATT	1139
Qy	1308	TGAATATAATGTGGAATGTCCTTCTTCCGGCACAATGAAACTTCCCTGCTCAATGTTGG	1367
Db	1140	TGAATACAATGTGGAACACCGCTGCCCAACAACATGAAACGACCGCTCTCAGTGGGAT	1199
Qy	1368	GAATTTCAAGTGGCATGGAATGAATGGTTGGGAGTGGCTGGAGCATATATCCAGAGTGG	1427
Db	1200	CAACTTCGAGTACAAGGGCATGACAGGCTGGGAGTGGCTGGTATCACATTTACACAGC	1259
Qy	1428	GGCTGGAGTTCTGACAATGATTACCTGATCTTAAACTTGCATATCCACGAGATTTAAACC	1487
Db	1260	TGCTGGAGCCTCAGACAATGACTTTCATGATTCTCACTCTGTTGTGCCAGGATTTAGACC	1319
Qy	1488	ACCAACATCTATGACTGGAGCCAC---GGGTCTCAACTAGGAAGAATAACATTTGCTTT	1544
Db	1320	TCCGACGTCGGTATGGCAGACACAGAGAATAAAGAGGTGGCCAGAAATACATTTGTCTT	1379
Qy	1545	TGACACCCCTCTGTTACGTGACTGTGTTTGTACTTTCATGTGGATATTATAGAAAAAG	1604
Db	1380	TGAGACCCCTCTGTTCTGTGNACTGTGAGCTCTACTTTCATGTGGGTGTGAATCTTAGAC	1439
Qy	1605	TACAAATGTGGTAGAATCGTGGGGTGGAAACCAAGAAAAACAAGCTTACACCCATATCAT	1664

Db	1440	CAACAC	TCTG	TGGAG	CGGTG	GGAAG	GGTTT	CCAAAG	GGCAAA	CAGTCT	TATAC	CTTAC	CTACAT	CAAT	1499
QY	1665	CTTCAAGA	ATGCA	ACTTTT	TACAT	TTTAC	ATG	GGC	CAATTC	CAGAG	AACTA	ATCAG	GGTCA	GA	1724
Db	1500	TGAGGAGA	CACTAC	CAAG	CTTCA	CTG	GGC	CTT	CAGAG	GACC	ACTTTT	CAT	GAG	GC	1559
QY	1725	TAATAGAC	GGTT	CATCA	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATG	1784
Db	1560	ANGCAGGA	AGTAC	ACCA	ATG	AGC	TTG	CCCA	AGAT	TACT	CCAT	CAAT	GTG	CCCA	1619
QY	1785	TGATGGGG	TGGC	TCCT	CAATG	CGC	TG	CGC	TG	CGC	TG	CGC	TG	CGC	1844
Db	1620	GAATGGGG	TGGC	TCTT	ACTG	CGC	TG	CGC	TG	CGC	TG	CGC	TG	CGC	1679
QY	1845	GTGTGTC	CCCTG	CCAG	CCACT	TACAT	TG	AGAA	GA	AAAC	CAAC	CAGT	GTG	CAAG	1904
Db	1680	CTGCAC	CTTGT	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	CTG	1739
QY	1905	TCCACCT	GTAC	ACTT	AC	CTG	TC	AT	AT	AT	AT	AT	AT	AT	1964
Db	1740	CCCCCT	TAAC	AAAT	TTG	AAAG	CC	CA	CCAG	CTT	ATG	GTG	TCC	AG	1799
QY	1965	CGGGCT	TGGG	AGT	AAAA	CAAT	CAG	AG	CA	CTT	CGG	TTG	CT	AT	2024
Db	1800	TGGTCC	AGG	GA	CAAC	AAAG	AT	TCC	ACT	CT	CTG	TG	CT	ACA	1859
QY	2025	CCATGAAA	AAAA	AAAT	CAG	ATTT	TG	CA	CT	AT	TG	CA	CT	AT	2084
Db	1860	AGCNA	CACT	CC	AC	AG	GA	CTTT	CA	CT	CA	CT	CT	CG	1919
QY	2085	ATTATG	AAAT	TG	CC	CC	AG	CTT	CA	CTT	CA	CTT	CA	CTT	2144
Db	1920	TCTTCT	TGG	AG	GC	CA	AG	CTT	CA	CTT	CA	CTT	CA	CTT	1979
QY	2145	CAGTTT	AT	TG	GG	CA	T	GAG	GG	AA	GA	T	GG	CT	2204
Db	1980	CAGTCT	CTG	TG	GA	AA	CC	AG	GGT	AG	AA	AT	TG	CT	2039
QY	2205	TACAGT	AAAA	CA	AAAT	AT	TG	CA	GG	CT	CAG	AT	AT	TG	2264
Db	2040	CGGAT	TCTG	AG	GGT	GAG	T	CA	GG	TT	-----	CT	CA	AA	2090
QY	2265	ATGCCAG	TC	AA	AT	T	AT	CT	CT	TG	AA	AT	AG	GG	2324
Db	2091	CTGCCAG	CA	CT	AT	CAT	CCCC	CA	AG	GGT	G	AG	GC	T	2150
QY	2325	ATCCAT	AT	CT	GG	CA	T	AT	CA	T	T	CA	GG	TC	2384
Db	2151	GCCTG	TC	AG	CC	T	AT	TG	GG	GT	G	ACA	CA	GA	2210
QY	2385	TAAAT	ATA	AA	GA	AT	TG	TT	CC	AG	TT	CC	AA	CA	2444
Db	2211	CAC	T	CC	CA	CT	TT	CC	AC	CT	GG	AG	TC	TT	2270
QY	2445	TTTAA	GT	CT	TT	C	AC	CA	CA	CT	CT	T	AT	TA	2504
Db	2271	TTATAG	G	TC	CA	AT	G	T	G	AC	CC	AG	T	CT	2330
QY	2505	GAG	GT	G	T	AT	CT	AA	TC	TG	AG	CA	GG	AT	2564
Db	2331	CAG	GT	CA	GT	CC	CA	GA	AA	AT	G	T	CC	T	2390
QY	2565	AG	T	AC	CT	G	T	G	T	AT	T	TC	T	TC	2624
Db	2391	TGG	GA	CT	T	G	T	CA	CT	TT	CC	AC	CT	TT	2450
QY	2625	GTG	T	AC	GG	AG	CA	T	GA	CT	TG	AG	GG	AG	2684
Db	2451	CTG	CT	C	AG	T	GG	CT	AT	CG	T	C	A	G	2510
QY	2685	CTT	G	T	AT	TG	TG	GA	AT	GA	CC	T	AA	T	2744

```
Db 2511 TACTTACGTGTGGCGAGAACCCCAAGCTATGCTCTGGTGGCAATTTCTCTCCCTGACGAGAG 2570
Qy 2745 GTTGGCAACCTGTGAACAGGTTTACATTTTGGCTGAAGGTG 2784
Db 2571 AGTCACCATCTGCAAAACCATAGATTCTCTGGCTGAAAGTG 2610

RESULT 8
AX127740
LOCUS AX127740
DEFINITION Sequence 15 from Patent WO0131003.
ACCESSION AX127740
VERSION AX127740.1 GI:14134387
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2733)
AUTHORS Delneste,Y., Magistrelli,G., Jeannin,P. and Bonnefoy,J.Y.
TITLE Cloning, expression and characterisation of a gene expressed in
tumour cells and involved in the regulation of the immune response
JOURNAL Patent: WO 0131003-A 15 03-MAY-2001;
PIERRE FABRE MEDICAMENT (FR)
FEATURES
Location/Qualifiers
1..2733
/organism="Homo sapiens"
/db_xref="GI:14134387"
/translation="MAEPGHSHLSARVGRTERIPRLRLLWAGTAFQVTOGTGP
ELHACKSEHYEYVACDSTGSRWVAVPHTPGLCTSLDPVKGTGECFSFCNAGEFLD
MKDQCKPCAEGRSLGTGIRFDEWDELPHGFASLNAWELDDSAESTGNCTSSKW
PRGDYASNTDECTALVAVNLKQSGTVNEFYXPDSSIIFEPFVONDOCPNADDS
RWMKTEGWBEHVELNRGNVLYWRTAESVWTKPKPVLVRIATIGVAYSECF
PKPGYADKQSSFCCKPANSYNGKETSCHOCDPKYSEKSSSCNVRPACTDKD
YFTTCDANGENTOLMYKWKAPKICSEDLEGAVKLPSGVKTHCPNCPNPGFFETNNS
QPCPCPYGNGSDCTRCPAETEPAGVEYKWNMTLPTNMETTIVLSCINFEYKGMTG
WEVAGDHLYTAAGSDNDFMLTLVPGFRPPQSVMDTENKEVARITVFETLCSVN
CELYPMVGNSTNTPVETWKGSKQSYVYIIEENTVTSFWAFORTTFHEASRKYT
NDVAKIYSINTVNVNGVASCPCALEASDVSGSCSCPAGYXIDRDSGTCHSCPNN
TILKHQPYGVQVQPCPGPTGNKNIHSLCYNDCFTSRNTPTRFNFNYSALANTVIL
AGPSTFSGKGLKFIHFHTLSLGNQGRMSVCTDNVDLRIPEGESGFSKSIATVCCQ
AVIIPPEVTGKAGVSSQPVSLADRLIGVTDMDLDTGSPAEILFHSLESLPDVIF
YRSDVTSQSSGRSTTIRVRCSPQKTPVGLLPLGTCSDGTCDGNCNPHFLWESAAC
PLCSVDYRAIVSSCVAGIQKTYVWREPKLCSGISLPEQRVTICKTIDFWLKV"
BASE COUNT 656 a 783 c 687 g 607 t
ORIGIN

Query Match 24.2%; Score 860.8; DB 6; Length 2733;
Best Local Similarity 60.0%; Pred. No. 3.5e-191;
Matches 1571; Conservative 0; Mismatches 1022; Indels 27; Gaps 7;

Qy 171 GCTTCCTCCTGCCAGGAGAAAGATTACATTTGTAATATACGGAATGTAGTAGAGTGG 230
Db 135 GCTTCATGCTGCAAGAGTCTGAGTACCACATATGATAGTACACGGCGTGTGACAGCAGG 194
Qy 231 CTCAGGTGGAGAGTTGCCATTCCTCAATCTGTCAGTGGAATCTGCTGGCCTGCCTGACCC 290
Db 195 TTCCAGGTGGAGGTGCGCGTCCGCATACACCCGGCGCTGTGCACAGCGCTGCCCTGACCC 254
Qy 291 AGTGAGAGCAAGAATGACATTTCTCCTGTGCTTGTGGAGAGTATCTAGAAATCAAGAA 350
Db 255 CGTCAAGGCGCAGGTGCTCTCTCTCTGCAACGCCGGGAGTTCTGGATATGAAGA 314
Qy 351 CCAGGTATGCAGTAAAGTGGTGAAGGCACCTATTCCTTTGGGCGAGTGGCATCAAAATTGA 410
Db 315 CCAGTATGTAAGCCATGCGCTGAGGGCGCGCTACTCCCTCGGCACAGGCATTCGGGTGA 374
Qy 411 TGAATGGGATGAATTCGCCGCGAGGATTTTCTAACATCGCAACATTCATGCACATCTGTGT 470
```

```
Db 375 TGAGTGGGATGAGCTGCCCCATGGCTTTGCCAGGCTCTCAGCCAAACATGGAGCTGGATGA 434
Qy 471 GGGCCCTTCTGACAGCAGGCGCAGCGGTGTAAACAACATCTTCTTGATCCCTCGTCGGA 530
Db 435 CAGTCTGCTGAGTCCA--CCGGGAACGTGACTTCGCTCCAAGTGGGTCCCGGGGGGA 491
Qy 531 CTACATAGAANTTAATCGTGTAGTACTGCACCGTGTCTTTTGATCTATGCTGTGCACCTTAA 590
Db 492 CTACATCGCCTCCAACACGGAGGAATGCACGCCACACTGATGTACGCGCTCAACCTGAA 551
Qy 591 GAACTCAGGCTATGCTTCTTTTGTAGTACAGTATGTGCACACACACATCTCTCTTTGAGTT 650
Db 552 GCAATCTGGCACCCTTAACTTCGAAATCTACTATCCAGACTCCAGCATCATCTTTGAGTT 611
Qy 651 CTTTATTCAAAATGATCAGTGCCAGGAGATGGACACCACTGACAAAGTGGGTAAAACT 710
Db 612 TTTCTGTCAGATGACCACTGCGCAGCCCACTGACATGACTC---CAGTGGATGAAGAC 668
Qy 711 TACAGACAATGGAGAAATGGGGCTCTCATTTCTGTAAATGCTGAATCAGGCACACAACTACT 770
Db 669 CACAGAGAA---AGGATGGAAATTCACAGTGTGGAGCTAAATCGAGGCAATAATGTCTT 725
Qy 771 CTACTGGAGAACTACAGCATCTCTTATGGGTCTTAAGGGGTCAAGCCTGTCTGCTAAA 830
Db 726 CTATTGGAGAACCAACAGCTCTCTCACTATGGACCAAGTACCACAGCCTGTCTGTGTAG 785
Qy 831 AAATATCAATTTGAAGGGGTGGCGTACACATCAGAAATGTTTTCTTCAAGCCAGGCAC 890
Db 786 AAACATTGCCATAACAGGGGTGGCGCTACACTTTCAGAAATGTTCCCTCGCAAACTTGGC 845
Qy 891 ATTACAGCAACAAACAGGTTCATTCACTGCCAGTGTGTGCCAGAAACACACTTATTCGA 950
Db 846 GTATGCAGACAGCAGGGCTCTCTTTCTGCAAACTTTGCCAGCAAACTCTCTATTCAAA 905
Qy 951 GAAAGGAGCCAAAGAAATGTATAGGTGTAA---ACAGCACTCTCAATTTTCAGGATCCAG 1007
Db 906 TAAAGAGCAAACTTCTGCCACCAGTGTGACCCCTGACAAATCTCAGAGAAAGGATCTTC 965
Qy 1008 TGAGTGTACAGAGCCCTTCCCTGTACCACAAAGCAATTTTCCAGATCCATPACCTCCATG 1067
Db 966 TTCTCTGTAAGTGGCGCCAGCTTGCACAGACAAAGATATTTCTACACACACACGCGCTG 1025
Qy 1068 TCATGAAGAAAGAAAGACACAGATTAATGTACAAGTGGATAGAGCCCAAAATCTGCCGGA 1127
Db 1026 CGATGCCAACGAGAGACACAACTCATGTACAAATGGCGCAAGCGAAATCTGTAGCGA 1085
Qy 1128 GGATCTCACAGATGCTATTAGATTGCCCTTCTCGAGAGAAAGAGGATTTGCCGCTTG 1187
Db 1086 GGACCTTGAGGGGCGAGTGAAGCTGCCTGCCTCTGTTGAGAGACCCACTGCCACCCCTG 1145
Qy 1188 CAACCTTGATTTTATAACAATGGATCATCTTTTGGCAATCCCTGTCTCTCTGGAACATTT 1247
Db 1146 CAACCCAGGCTTCTTCAAAACCAACAGCAGCTTGCAGGCGCTGCCCATATGTGCTCCTA 1205
Qy 1248 TTCAGATGAACCAAGAATGTAGACCATGTCCACAGGAGGAGCCCTGCACTTTGGCTT 1307
Db 1206 CTCCAATGG---CTCAGACTGTACCCGCTGGCCTCGAGGACTGAACCTGTCTGGGATTT 1262
Qy 1308 TGAATATAAATGGTGGAAATGTCTTCTTCCGCAACATGAAACCTTCTGCTTCAATGTTGG 1367
Db 1263 TGAATACAAATGTTGGAACACGCTGCCCAACAAACATGGAAGACCGTCTCAGTGGGAT 1322
Qy 1368 GAAATTCAAAGTCGATGGAATGAATGTTGGAGGTGGTGGAGATCATATCCAGAGTGG 1427
Db 1323 CAACTTCGAGTACAAAGGCGATGACAGGCTGGGAGGTGGCTGTGATCATATTTACACAGC 1382
Qy 1428 GGTGAGGTCTGACAAATGATTACCTGATCTTAACCTTGCATATCCAGGAGATTTAAACC 1487
Db 1383 TCGTGGAGCCTCAGACAAATGACTTTCATGTTCTCACTCTGTTGTGGCAGGATTTAGAG 1442
Qy 1488 ACCAACAATCTATGACTGGAGCCAC---GGGTCTCTGAACCTAGGAAGAATAAATTTGTCTT 1544
```

Db	1443	TC	CGCAGTCGGTGCATGGCAGACACAGAGAAATAAGAGGTGGCCAGAAATCACATTTGTGCTTT	1506
Qy	1545	TG	AGACCCCTCTGTTTACGCTGACTGTGTTTGTGTACTTTCATGGTGGATAATTAATAGAAAAG	1604
Db	1503	TG	AGACCCCTCTGTTTCTGTGAACGTGTGAGCTCTACTTTCATGTGGTGTGAATTCCTAGGAC	1562
Qy	1605	TACA	ATGTGGTTAGATCTGTGGGTGGAAACCAAGAAACAAAGCTTTACACCCATATCAT	1664
Db	1563	CA	ACACTCTCTGTGGAGACGTGGAAGAGTTTCCAAGGCAAAACAGCTTATACCTACATCAT	1622
Qy	1665	CTTC	AGAATGCAACTTTTACATTTTACATGGGCATTCAGAGAACTAATCAGGGTCAAGA	1724
Db	1623	TG	AGGAGAACACTACCAACGAGCTTCACCTGGCCCTTTTCAGAGGACCACTTTTCATGAGGC	1682
Qy	1725	TAAT	PAGACGGTTTCATCAATGACATGGTGAAGATTTATTTCTATACAGCCACCTAATGTCAGT	1784
Db	1683	AGC	AGAAGTACACCAATGACGTGGCCAAAGATCTACTCCATCAATGTCACCAATGTTAT	1742
Qy	1785	TGAT	GGGGTGGCTCTCATCGCGTGGCTGTGCCCTTCGGTTCTTGAACAGTGGGGTTCAATC	1844
Db	1743	GAAT	GGCGTCCCTACTGCGGTCCCTGTGCCCTGAAGGCCTCTGATGTGGGCTCCTC	1802
Qy	1845	GTGT	GTCCCTGCCCTCCAGGCCACTACATTCAGAAAAGAAACCAACAGTGCAGAGGAATG	1904
Db	1803	CTG	CACCTCTGTGCTGTGGTTACTATATTAACCGAGATTTCAGGAACCTGCCACTCCTG	1862
Qy	1905	TCC	ACCTGCACACCTACTGCTCCATACATCAGGTCTATGGCAAGAGGCTGTATTTCCATG	1964
Db	1863	CCCC	CTTAACACAATTCGTGAAGCCCAACAGCCCTTATGGTCCAGGCCCTGTGTGCCCTG	1922
Qy	1965	CGG	CCCTGGGAGTAAACAACATCAGACCATTCGGTTTGCTATAGTGAAGCTCTTTTCTTA	2024
Db	1923	TGGT	TCCAGGGACCAAGAACAAGATCCACTCTCTGTGCTACAATGATTCACACTTCTC	1982
Qy	2025	CCAT	GAAGAAAGAAATCAGATTTTGACATACATCTTACATCTTAGCAACCTCAGCAGTGTGGGCTC	2084
Db	1983	ACG	CAACCTCCAAACGAGGACTTTCAACTTACAACCTTCTCCGCTTTGGCAAAACACCGTCAC	2042
Qy	2085	ATTA	ATGAATGGCCCCAGCTTCACCTCCAAAGGAACAATACTTCCATTTCTTCCAATAT	2144
Db	2043	TC	TGCTGGAGGCCAAGCTTCACTTCCAAGGGTTGAATYACTTCCATCATCTTTACCCCT	2102
Qy	2145	CAGT	TTATGTGGCGATGAGGGGAAGAAGTCTCTGTATCCAAACAATATAACAGACTT	2204
Db	2103	CAGT	CTCTGTGGAAACAGGGTAGGAAAATGTCTGTGTGCACCGACAATGTCACTGACCT	2162
Qy	2205	TAC	AGTAAAGAAATAGTGCAGGCTCAGATGATACACAATTTGGTAGGGGCAATTTGT	2264
Db	2163	CCG	GAATTCCTGAGGGTGAGTCAGGGTT-----CTCCAAATCTATCACAGCCTACGT	2213
Qy	2265	ATG	CAGCTCAACAATTTATTCCTTCTCAAAAGTAAGGGTTTCCGAGCAGCTTTATCATCACA	2324
Db	2214	CTG	CCGCGAGTCATCATCCCCCAGAGGTGACAGGCTACAGGCCGGGGTTTCTCTACA	2273
Qy	2325	ATC	CATCATTTCTGGCAGATACATTCATAGGAGTCAAGTTGAAACACCAATTAAGAAAATAT	2384
Db	2274	GCCT	GTCAAGCCTTGCTGATCGACTTATGGGGTGACAACAGATATGACTCTGGATGGAAT	2333
Qy	2385	TAAT	TATAAAGAAGATATGTTCCCGTTCCACAGCCCAATACCAGATGTGCATTTTCTT	2444
Db	2334	CAC	TCCCCAGCTGAACTTTTCCACCTGGAGTCTTGGGAATACCGGACGTGATCTTCTT	2393
Qy	2445	TTAT	AAAGTCTTCTACAGCAACAACATCTTGTATTAATGGCCGATCAACTGCTGTGAAAT	2504
Db	2394	TTAT	AGGTCCAATGATGTGACCCAGTCTCTGCAGTCTTGGGAGATCAACACCANTCCGGGT	2453
Qy	2505	GAG	GTAAATCCTACTAAATCTGGAGCAGGAGTGATTTTCAGTCCCCAGCAAGTCCCGCAGC	2564
Db	2454	CAG	GTGAGTCCACAGAAAACTGTCCCTGGAAGTTTGCTGTGCCAGGAACGTGCTCAGA	2513
Qy	2565	AGG	TACCTGTGATGGGTGACGTTTCTATTTCTGTGGGAGAGTGTGTAAGCTTGGCCCTCT	2624
Db	2514	TG	GACCTGTGATGGGTGCAACTTCCACTTCTCTGTGGAGAGCGCGCTGCTGCGCGCT	2573

Qy	2625	GTGTACGAGCATGACTTCCATGATAGATTGAGGGAGCCTGCAAGAGAGAGATTTCACGAAAC	2684
Db	2574	CTGCTCAGTGGCTGACTACCGTGTCTATCTGTCAGCAGCTGTGCTGGATCCAGAAGAC	2633
Qy	2685	CTTGATGCTGTGGAAATGAACCTAAATGGTGCATTAAAGGAATTTCTTTGGCTGAGAAAAA	2744
Db	2634	TACTTACGTGTGGGAGAACCCACAGCTACTCTGTGGTGCAATTTCTCTGCTGAGCAGAG	2693
Qy	2745	GTGTGCCAACCTGTGAACGGTTGACTTTTGGCTGAAGGTG	2784
Db	2694	AGTCACCATCTGCCAAACACCATAGATTTTCTGCTGAAAGTG	2733

RESULT 9

AK057647

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

BASE COUNT

ORIGIN

2895 bp

linear

PRI 01-AUG-2002

Homo sapiens cDNA FLJ33085 fis, clone TRACH2000420, weakly similar to MAJOR SURFACE-LABELLED TROPHOZOITE ANTIGEN PRECURSOR.

AK057647

AK057647.1

GI:16533406

oligo capping: fis (full insert sequence).

Homo sapiens trachea cDNA to mRNA, clone_lib:TRACH2

clone:TRACH2000420.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

1

Suzuki,O., Sasaki,N., Aotsuka,S., Shoji,T., Ichihara,T., Shiohata,N., Matsumoto,K., Hirano,M., Sano,S., Nomura,R., Yoshikawa,Y., Matsumura,Y., Woriya,S., Chiba,E., Momiyama,H., Onogawa,S., Kaeriyama,S., Satoh,N., Matsunawa,H., Takahashi,E., Kataoka,K., Kuga,N., Kuroda,A., Satoh,I., Kamata,K., Takami,S., Terashima,Y., Watanabe,M., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Ota,T., Wakamatsu,A., Ishii,S., Yamanoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Wadatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.

NEDO human cDNA sequencing project

2 (bases 1 to 2895)

Isogai,T., Otsuki,T. and Sugiyama,T.

Direct Submission

Submitted (24-OCT-2001)

Takao Isogai, Helix Research Institute, Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp. Tel:81-438-52-3975, Fax:81-438-52-3986)

NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan: cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.

Location/Qualifiers

1..2895

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="TRACH2000420"

/tissue_type="trachea"

/clone_lib="TRACH2"

/note="cloning vector: pME18SFL3"

692 a

825 c

714 g

664 t

20.6%

Score 732.8;

DB 9;

Length 2895;

Best Local Similarity 57.3%;

Pred. No. 4.3e-161;

Matches 1575;

Conservative 0;

Mismatches 1002;

Indels 171;

Gaps 7;

171

GCTTCCTCCTCCAGGAGAAAGATTATCACTTTGAATATACGGAATGTATGATGACGTGG

230

Db 214 GCTTCATGCCGTGCAAGAGTCTGATACCACATATGAGTACACGGGTGTGACAGCAGCGG 273
Qy 231 CTCAGGTGGAGAGTTGCCATTTCCTCAAAATCTGCAGTGCAGTCTCTGGCCTCGCTGACCC 290
Db 274 TTCAGGTGGAGGTGCGCGTGCAGCATACCCGGGCGTGTGCACCGCTCGCTGACCC 333
Qy 291 AGTAGAGGAAAGAAATGCATCTCTCTCTGCTGTCTGGAGAGTATCTAGAAATGAAGAA 350
Db 334 GCTCAAGGCAACCGAGTCTCTCTCTGCAAGCGCGGAGTTCTCTGGATATGAAGGA 393
Qy 351 CCAGTATGCAATAGTGTGTGAAGGCACCTATTCCTTTGGCAGTGCATCAAAATTTGA 410
Db 394 CCAGTATGTAAGCCATCGCTGAGGCGGTACTCTCCCTCGGCACAGCATCTGGTTTGA 453
Qy 411 TGAATGGGATGAATTTGCCGGCAGGATTTCTTAACATCCAAATCTTCATGGACACTGTGGT 470
Db 454 TGAGTGGGATGAGTGGCCCATGGCTTTGCCAGCGCTCTCAGCCAACTGGAGCTGGATGA 513
Qy 471 GGGCCCTTCTGACAGCAGGCCAGCGCTGTAAACACTCTCTTGGATCCCTCGTGGAA 530
Db 514 CAGTCTGCTGAGTCCA---CGGGAATGTACTTCGCTTCCAAAGTGGTTCGCCGGGGCGA 570
Qy 531 CTACATAGAATCTAATCTGTGATGACTGCACGGTGTCTTTGATCTATGCTGTGCACCTTAA 590
Db 571 CTACATCGCTCCACACGGGAGATGCACAGCCACTGATGTACGCGGTCAACCTGNA 630
Qy 591 GAAGTCAGCTATGCTCTTTGAGTACAGATGTGTCGACAAACATCTCTCTTTGAGTT 650
Db 631 GCAATCTGGCAGCGTTAACTTCGAATCTACTATTCAGACTCCAGCATCATCTTTGAGTT 690
Qy 651 CTTTATTCAAATGATCAGTGCAGGAGATGGACACCACCCTGACAAAGTGGTAAACT 710
Db 691 TTTCTGTTCAAGATGACAGTGCAGCCCAATGCAGATGACTC---CAGGTGGTGAAGAC 747
Qy 711 TACAGCAATGAGAGTGGGCTCTCATTTCTGTAATGCTGTAATCAGGCACCAAAACATACT 770
Db 748 CACAGAA---AGGTGGGAATTCACAGTGTGGAGCTAATCGAGGCAATATGCTCT 804
Qy 771 CTACTGGAGAACTACAGGATCCTTATGGGTTCTAAGGGGTCAAGCCTGTGCTGGTAAA 830
Db 805 CTATTGGAGAACACAGCGCTTCTAGTATGGACCAAAAGTACCACAGCCTGTGCTGTGAG 864
Qy 831 AAATACCAATTTGAAGGGTGGCGTACACATCAGAAATGTTTCTCTGCAAGCCAGGCAC 890
Db 865 AAACATGGCCATACACAGAA-----885
Qy 891 ATTACAGCAAAACCAGGTTCATTCAACTGCCAGGTGTGCCAGAAACACCTATTCTGA 950
Db 886-----885
Qy 951 GAAAGGAGCAAGAAATGTATAGGTGTAAAGACGACTCTCAATTTTCAGGATCCAGTGA 1010
Db 886-----AGGATCTCTTC 897
Qy 1011 GTGTACAGAGCGCTCCCTGTACCAAAAGACTATTTCCAGATCCATPCTCCATGTGA 1070
Db 898 CTGTACGTGGCGCCAGCTGTGCACAGACAAAGATTTATTTCTACACACACCGCCTGCGA 957
Qy 1071 TGAAGAAGAAAGACACAGATATGTACAAAGTGGATAGAGAGCCCAAAATCTGCCGGGAGGA 1130
Db 958 TGCAACGGAGAGACACAACACTCATGTACAAATGGGCAAGCGGAAATCTGTAGCGAGGA 1017
Qy 1131 TCTCAGATGCTATTAGATTGGCCCTTCTGGAGAGAAAGAGGATGTCCGCCCTTGCAA 1190
Db 1018 CTTTGGGGGGGAGTGAAGCTGCCCTCTGTTGTGAAGACCACTGCCACCCCTGCAA 1077
Qy 1191 CCTGGATTTTATAAAGATGGATCATCTTTCTGCCATCCCTGCTCTCTGGAACATTTTC 1250
Db 1078 CCCAGCTTCTTCAAAACCAACACAGCAGCTGCCAGCCCTGCCATATGGTTCTCTACTC 1137
Qy 1251 AGATGAACCAAGAAATAGACCATGTCCAGCAGGAAGCGCTGTGACCTTGGCTTTGA 1310
Db 1138 CAATGG---CTCAGACTGTACCGCTGCCCTGCAAGGAGTGAACCTGCTGGGATTTGA 1194

Qy 1311 ATATAAATGGTGGAAATGTCTCTTCTCGCAACATGAAAACTTCTCTGCTTCATATGTTGGAA 1370
Db 1195 ATACAAATGGTGGAAACACGCTGCCACAAACATGAAACGACCGTCTCTCAGTGGGATCAA 1254
Qy 1371 TTCAAAGTGCATGAAATGGTGGAGGTGGCTGGAGATCATATCCAGAGTGGGGC 1430
Db 1255 CTTGAGTACAAAGGCATGACAGGCTGGAGGTGGCTGGTATCACAATTTACACAGCTGC 1314
Qy 1431 TGGAGTTTCTGACAAATGATACCTGATCTTAACTTGCATATCCAGGATTTAAACACC 1490
Db 1315 TGGAGCTTCAGACAATGACTTTCATGATCTCCTCTGTTGTGCCAGGATTTAGACCTCC 1374
Qy 1491 AACATCTATGACTGGAGCCAC---GGGTCTTGAACCTAGAGAAATAAGATTTGCTTTGA 1547
Db 1375 GCAGTCGGTGTATGCGACACACAGAGAATAAAGAGTGGCCAGAAATCAATTTGCTTTGA 1434
Qy 1548 GACCTCTGTTGACGTGACTGTGTTTCTTACTTCTATGCTGGATATTAATAGAAAAAGTAC 1607
Db 1435 GACCTCTGTTCTGTGAACCTGTGAGCTCTACTTCTATGGTGGTGTGAATTTCTAGGCCAA 1494
Qy 1608 AAATGTTAGAAATGGTGGGTGGAAACCAAAAGAAACAAAGCTTTACACCCATATCATCTT 1667
Db 1495 CACTCCTGTGGAGAGCTGGAAAGGTTCCAAAGGCAACAGCTCTATACCTACATCTTGA 1554
Qy 1668 CAAGANTCCAATTTTACATTTTACATGGGCATTTCCAGAGAACTAATCAGGGTCAAGATAA 1727
Db 1555 GGAGAACACTACACAGAGCTTCACCTGGGCTTCCAGAGGACCACCTTTTCTATGAGGCAAG 1614
Qy 1728 TAGCGGTTTCATCAATGACATGGTGAAGATTTATTTCTATACAGCCACTAATTCAGTTGA 1787
Db 1615 CAGGAAGTACCAATGAGTGTGCCAAGATCTACTCCATCAGTGTCCACAAATGTTATGAA 1674
Qy 1788 TGGGTGGCGTCTCATGCCGTGCTGTGCCCTCGGTCTGAAACAGTGGGTTCATCGTG 1847
Db 1675 TGGGTGGCGTCTCTACTGCCGTCCCTGTGCCCTAGAAAGCTCTGATGTGGGCTCCTCTG 1734
Qy 1848 TGTCCCTTCCCTCCAGGCCACTACATTTGAGNAAGAACCAACACAGTGCAGAGGATGCC 1907
Db 1735 CACCTCTTGTCTGCTGCTGGTACTATATTGACGAGATTTCAAGAACCTGCCACTCCTGCC 1794
Qy 1908 AGCTGACACCTACCTCTCCATACATCAGCTCTATGGCAAGAGGCTGTGTTTCCATGCGG 1967
Db 1795 CCTAACAACATTTCTGAAAGCCACAGCCTTATGGTGTGCCAGGCTGTGTGCCCTGTGG 1854
Qy 1968 GCCTGGGAGTAAAAACAATCAGGACCATTTGGTCTATAGTGAAGTCTTTCTACCA 2027
Db 1855 TCCAGGGACCAAGAACAAACAGATCCACTCTCTGTGCTACAAATGATTCACCTTCTCAG 1914
Qy 2028 TCAAAAAGAAATCAGATTTTTCAGTATGACTTTTAGCAACCTCAGCAGTGTGGGCTCATTT 2087
Db 1915 CAACACTCAACCCAGACTTTCACTACACTTCTCCGCTTTGGCAACACACGCTCACCT 1974
Qy 2088 AATGAATGGCCCGCAGCTTCACCTCCAAAAGAAACAAATACTTTCCATTTCTCAATATCAG 2147
Db 1975 TGCTGGAGGGCCAAAGCTTCACCTTCCAAAAGGTTGAAATACTTCCATCTACTTTACCTCAG 2034
Qy 2148 TTTATGTTGGCATGAGGGGAGAAAGATGGCTCTCTGTACCAACAATATTAACAGACTTTAC 2207
Db 2035 TCTCTGTGAAACACCCAGGTAGGAAATGCTGTGTGACCCGACAAATGTACACTGACCTCGG 2094
Qy 2208 AGTAAAGAAATAGTGGCAGGGTCAAGTGTATACAAATTTTGGTAGGGGCAATTTGTATG 2267
Db 2095 GATTCCTGAGGTGAGTCAAGG-----TTCTCCAATCTATCACAGCCTACGTCG 2145
Qy 2268 CCAGTCAACAATTTATTTCTTCTGAAAGTAAAGGTTTCCGAGCAGGCTTTATCATCACAATC 2327
Db 2146 CCAGGAGTTCATCATCCCGCCAGAGGTACAGGCTACAAAGCGCGGGGTTTCTCTCACAGCC 2205
Qy 2328 CATCATTTGGGAGATACATTTATAGGAGTCAAGTGTGAACCCACATTTCAAAAATATTTAA 2387
Db 2206 TGTAGCCCTTGTCTGACGACTTATTTGGGGTGACACAGATATGACTCTCGGATGGAATCAC 2265

QY	2388	TATAAAGAAGATATGTTCCAGTTCACAAAGCCAAATACCAGATGTGCAATTCCTTTTA	2447
Db	2266	CTCCAGCTGAACATTTTCCACCTGGAGTCTTGGGAATACCGGACGTGATCTCTTTTA	2325
QY	2448	TAAGTCTTCTACAGCACACATCTTGTATTATGTCGGCATCAACTGCTGTGAATGAG	2507
Db	2326	TAGTCCAAATGATGTACCCAGTCTTCGACGTTCTGGAGATCAACCACTCCGCGTCAG	2385
QY	2508	GTGTAATCTTACTAAATCTGGAGCAGAGTGATTTTCAGTCCCGAGCAAGTGCACAGG	2567
Db	2386	GTGAGTCCACAGAAACTGTCCTGGAAGTTTGCTGCTCCAGGACGTCCTCAGATGG	2445
QY	2568	TACCTGTGATGGGTGACGTTCTATTTCCTGCGGAGAGTCTCAAGCTTTCCTCTGTG	2627
Db	2446	GACCTGTGATGGGTGCACTTCCATCTCCTGTGGAGAGCGCGCTGCTTGCCTGCTG	2505
QY	2628	TACGGACATGACTTCCATCAGATGAGGAGGCTGCAAGAGAGGATTCAGGAACCTT	2687
Db	2506	CTCAGTGGGTGACTACCATCTATCGTCAGCAGCTGTGTGGCTGGGATCCAGAAGACTAC	2565
QY	2688	GTATGTGTGGAATGAACCTAAATGGTGCATTAAAGGAATTTCTTGGCTCAGAAAAAGTT	2747
Db	2566	TTAGTGTGGCGAGAACCCCAAGCTATGCTCTGGTGGCATTTCTGCTCAGCAGAGT	2625
QY	2748	GGCAACCTGTGAACGGTGTGACATTTTGGCTGAAGGTGGGAGCGGTGTGGAGCTTTTAC	2807
Db	2626	CACCATCTGCAAAACCATAGATTTTGGCTGAAAGTGGGCATCTCTCAGCGACCTGTAC	2685
QY	2808	TGCGTTTCTGCTGGTGGCTGACCTGCTACTTCTGCGGAGAGAAATCAAAACTGGNATA	2867
Db	2686	TGCACTCCTCTCACCCTCTGACCTGCTACTTTTGGGAAAGAAATCAAAACTAGAGTA	2745
QY	2868	CAAAATTTCCAAGTTAGTATGACGACTAACTCAAAAGAGTGTGAACCT	2915
Db	2746	CAAGTACTCCAGCTGTGATGATGCTACTCTCAAGGATTTGGCCT	2793
RESULT	10		
LOCUS	BC029010	2188 bp	linear ROD 07-AUG-2002
DEFINITION	Mus musculus, clone IMAGE:5361249, mRNA		
ACCESSION	BC029010		
VERSION	BC029010.1	GI:20810059	
KEYWORDS	house mouse.		
SOURCE	Mus musculus		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 2188)		
AUTHORS	Strausberg, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (01-MAY-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov		
COMMENT	Contact: MGC help desk		
	Email: cgapbs-r@mail.nih.gov		
	Tissue Procurement: The Cepko Laboratory		
	cDNA Library Preparation: Life Technologies, Inc.		
	DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
	DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center		
	Center code: BCM-HGSC		
	Web site: http://www.hgsc.bcm.tmc.edu/cdna/		
	Contact: amg@bcm.tmc.edu		
	Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M., Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.		
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov		
	Series: IRAK, Plate: 54 Row: O Column: 14		

FEATURES		Location/Qualifiers
source	1..2188	/organism="Mus musculus"
		/db_xref="taxon:10090"
		/clone="IMAGE:5361249"
		/tissue_type="Eye, retina, mouse strain C57BL/6"
		/clone_lib="NIH_MGC_94"
		/lab_host="DH10B"
		/note="Vector: pCMV-SPORT6"
CDS	<1..720	
	/codon_start=1	
	/product="Unknown (protein for IMAGE:5361249)"	
	/protein_id="AAH29010.1"	
	/db_xref="GI:20810060"	
	/translation="DTALQNVNIKEDMPVSPQVDVHFVYKSSATTTSCINGRSTA VKMRNPMRPGAGVISVPSKPGAGTCDGCTEYFLWESAECPLCKEHEIEKKKR GLQEILYVNEPKWCIKIGISLPEKILSTCTETDFWLKVGAGFATVALLVALTCYFW KKNOKLEYKYSKLVMTTNSKECELPADSCAIMEGEDNEEDVYVSNKSLGLKLKLSLA TREKDDHFESVQLKSRCPNI"	
BASE COUNT	695 a 360 c 453 g 680 t	
ORIGIN		
Query Match	20.0%;	Score 711.6; DB 10; Length 2188;
Best Local Similarity	80.8%;	Pred. No. 3.8e-156;
Matches	921; Conservative	0; Mismatches 174; Indels 45; Gaps 6;
QY	2365	GAAACCACTTGAATAATATAATAAAGAGATATGTTCCAGTTCACCAAGCCAA: 2424
Db	1	GACACTGCTTGAACAGTTAATATCAAGGAGATATGTTCCAGTGTCCAGAGCAA 60
QY	2425	ATACCAGATGTCATTTTATAAGTCTCTACACACACACATCTTGTATTATGCG 2484
Db	61	GTACCAGATGTTCTTTTATAAGTCTTCCACACCCACACATCTGTATTATGGA 120
QY	2485	CGATCAACTGCTGGAATAAGGTTAATCTACTAAATCTGGAGCAGGAGTATTCA 2544
Db	121	CGATCGACTGCTGACAGATGAGTGTATCCCATGAGACCTGGTGCAGGTGATATCA 180
QY	2545	GTCCCCAGCAAGTGCACGAGTACCTGTGATGGGTGTACGTTCTATTTCCTGTGGAG 2604
Db	181	GTCCCCAGCAAGTGCACGAGTACCTGTGATGGTGTACCTTCTTCTTCTTCTTCTTCT 240
QY	2605	AGTGCTGAAGCTTGCCTCTGTGTACGAGCATGACCTTCATGAGATGAGGGAGCCTGC 2664
Db	241	AGTGCAAGCTTGCCTCTGTGCAAAAGACATGACTTCCATGAGATCGAGAAAAAAA 300
QY	2665	AAGAGAGGATTTTCAGGAAACCTTGTATGTGTGAATGAACCTAAATGTGCTATTAAAGGA 2724
Db	301	AAAGAGGGCTTCAGAAATATATATATGATGGAATGAACCTAAGTGGTGCATTAAGGA 360
QY	2725	ATTTCCTTGGCTGAGAAAAAGTTGGCAACCTGTGAAACGGTTGACCTTTTGGCTGAAGGTG 2784
Db	361	ATTTCCTTGGCCGAAAGAGTGTGTCAACCTGCGAAACCTGTGACTTTTGGCTCAAAAGTG 420
QY	2785	GGACCGGTGTGGAGACTTTTACTGCGGTTTGTGCTGGTCTGACCTGCTACTTCTTGG 2844
Db	421	GGAGCTGGTGTGGCGCGTTTACAGCGGTTTGTGTTGGTGGCTTTTAAATGATGCTACTTGTG 480
QY	2845	AAAAGAATCAAAACCTGGAATACAAATATTTCCAAAGTTAGTAATGACGACTTAACCTCAAAA 2904
Db	481	AAAAGAATCAAAAGCTGGAAATACAAATATTTCTAAATAGTAATGACGACTTAACCTCAAAA 540
QY	2905	GAGTGTGAATCCCGCTGCAGACAGTGTGCTATCATGGAAGGAGAGATTAATGAAGAG 2964
Db	541	GAGTGTGAATCCCGCTGCAGACAGTGTGCTATCATGGAAGGAGAGATTAATGAAGAG 600
QY	2965	GAGTGTGATATTCATTAACAGTCACTACTAGGAAAACTCAAAATCTTTGGCAACCAAG 3024
Db	601	GATGTTGTATATTCACCAACACAGTCACTACTAGGAAAACTTAAGTCTCTTGGCCACCAAG 660

This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

Qy	1679	CTTTTACATTTACATGGCATTCCAGAGACTAATACAGGTCACAGATTAATAGACGGTTCA	1738
Db	794	CCAGAGCTTCACTCGGGCTTCCAGAGACACATTTTCATGAGGCAAGCAGGAAGTACA	853
Qy	1739	TCAATGACATGGTGAAGATTTATCTATCACAGGCACATAATGCAGTTGATGGGTGGCGT	1798
Db	854	CCAATGAGTGGCAAGACTACTCCATCAATGTCACCAATGTTATGANTGGCGTGGCCT	913
Qy	1799	CCTCATGCCCTGCTGTGCCCTCGGTTCTGAAACAGTCGGGTTCACTGTTGTCCTCCCTGCC	1858
Db	914	CCTACTGCCCTGCTGTGCCCTAGAAGCCTCTGATGTGGGCTCCTCTGCACTCTTGTC	973
Qy	1859	CTCCAGGCCACTACATTTGAGAAGAAACCAACCACTGCAAGGAATGTCCACCTGCACACT	1918
Db	974	CTGCTGGTTACTATTTGACGAGATTCAAGAACCTGCGCACTCTGCCCCCTTAACACAA	1033
Qy	1919	ACCTGTCCATACATCAGGTTCTATGSCAAAGAGCGTTGATTCCATGGGGCTGGAGTA	1978
Db	1034	TTCTGAAAGCCCACGACCTTATGTTGTGCCAGGCGCTGTGCCCTGTGTCAGGACCA	1093
Qy	1979	AAACAATCAGGACCATTCGGTTTGCTATGTAGTACTGCTTTTCTACCATGAAGAGAA	2038
Db	1094	AGAACAAGATGCCACTCTCTGTGCTACAAATGATGTACACTTCTCAGCAACACTCAA	1153
Qy	2039	ATCAGATTTTGCACTATGACTTTAGCAACCTCAGCAGTGTGGGCTCATTAATGAATGGCC	2098
Db	1154	CCAGGACTTTCACATACACTTCTCGGCTTTGGCAACACCGTCACCTCTGCTGGAGGGC	1213
Qy	2099	CCAGCTTTCACCTCCAAAGGAACAAATACTTCCATTTCTTCAATATCAGTTTATGTGGC	2158
Db	1214	CAAGCTTTCACCTCCAAAGGTTGAAATACTTCCATCACTTTACCCTCAGTCTCTGTGAA	1273
Qy	2159	ATGAGGGAAGAGATGGCTCTCTGTACCAACAATTAACAGACTTTFACAGTAAAGAAA	2218
Db	1274	ACCAGGTAGGAANAATGTCTGTGTGCACCGCAATGTCACTAGCACTCGGATTCCTGAGG	1333
Qy	2219	TAGTGGCAGGGTCAGATGATTACACAAATTTGGTAGGGCAATTTGATGCCAGTCAACAA	2278
Db	1334	GTGAGTCAGGG-----TTCTCCAATCTATCACAGCCTACGCTGCCAGGCACTCA	1384
Qy	2279	TTATTCCTCTGAAAGTAAGGTTTCCGAGCAGCCTTATCATCAATATCCATTCATTTCTGG	2338
Db	1385	TCATCCCCCAGAGGTGACAGGCTACAAGGCGGGGTTTCTCACAGCCTGTGACGCTTG	1444
Qy	2339	CAGATACATTCATAGGAGTACAGTTTGAACACCAATTCGAAANAATTAATATAAGAGAAG	2398
Db	1445	CTGATCAGACTATTGGGGTGAACACAGATATGACTCTGGATGGAATCACCTCCCCAGCTG	1504
Qy	2399	ATATGTTCCAGTTTCCAACAAGCAAATACCAAGATGTGCATTTCTTTTAAAGTCTTCTA	2458
Db	1505	AACCTTTCCACCTGGAGTCTTGGGNAATACCGAGCTGATCTCTTTTATAGTCCCATG	1564
Qy	2459	CAGCAACAACATCTTGTTAATATGGCGGATCAACTGTGCTGTAAGATGAGGTAACTCTA	2518
Db	1565	ATGTGACCCAGTCTCTCAGTTCTGGGAGATCAACCAACCATCCGCGTCAGGTGCACTCCAC	1624
Qy	2519	CTAAATCTGGAGCAGGAGTGATTTCAAGTCCCCAGCAAGTCCCCAGCAGGTACCTGTGATG	2578
Db	1625	AGAAACTGTCCTGGAAGTTTGCTGTGCCAGGAACGTCTCAGATGGGACCTGTGATG	1684
Qy	2579	GGTGTAGTTTCTATTTTCCTGTGGGAGTGTGTAAGCTTTGCCCTCTGTGTACGAGCAGT	2638
Db	1685	GCTCAACTTCCACTTCTGTGTGGAGAGCGGCTGTGTCGCCCTCTGCTCAGTGGCTG	1744
Qy	2639	ACTTCATGAGATTTGAGGAGCCTGCAAGAGAGGATTTTCAGGAACCTTGATGTGTGGA	2698
Db	1745	ACTACCATGCTATCTGTCAGCAGCTGTGTGGCTGGGATCCAGAACACTTACCTGTGTGC	1804
Qy	2699	ATGAACCTAAATGGTCCATTAAGGAATTTCTTCCTGTGAAAAAGTTTGGCAACCTGTG	2758
Db	1805	GAGAACCAAGCTATGCTCTGGTGGCATTTCTGCTCAGCAGAGAGTCAACCATCTGCA	1864

QY	2759	AAACGGTTGACTTTTGGCTGGAAGTGGAGCGGTTGGAGAGCTTTTACTGCCCTTTTGC	2818
Db	1865	AAACCATAGATTCTTGGCTGGAAGTGGGCACTCTTGCAGGACACCTGTACTGCCATCTCG	1924
QY	2819	TGCTGGCTCTGACCTTCTCTCGAAAAAGAAATCAAAAACTGGAATACAAAATATTC	2878
Db	1925	TCAACCTCTTGACCTGCTTCTTGGAAAAAGAAATCAAAAACTAGATGACAACTTCCA	1984
QY	2879	AGTTTGTATGACGACTAAGCTCAAAAGAGTGTGAACCTCCGGCTGCAGACAGTTTGTCTA	2938
Db	1985	ACCTGGTGATGAATGCTACTCTCAAGGACTGTGACCTGCCAGCAGCTGACAGCTGCGCA	2044
QY	2939	TCATGGAAGGAGAAGCAATGAAGAGGAGTTGTATATTTCCAATAAAACAGTCACTACTAG	2998
Db	2045	TCATGGAAGGCGAGATGTAGAGGAGCAGCTCATCTTTTACCAGCAAGATCACTCTTTG	2104
QY	2999	GAAACTCAAACTTTTGGCAACCAAGAGGAAAAAGAACACTTTTGAATCTGTTCAACTGA	3058
Db	2105	GGAAGATCAAAATCATTTACCTCCAAGAGGACTCCTGATGGATTTGACTCAGTGGCGCTGA	2164
QY	3059	AAACCTC 3065	
Db	2165	AGACATC 2171	
RESULT	12		
LOCUS	AK026832	2129 bp mRNA linear	PRI 29-SEP-2000
DEFINITION	Homo sapiens cDNA: FLJ23179 fis, clone LNG10890.		
ACCESSION	AK026832		
VERSION	AK026832.1 GI:104399779		
KEYWORDS	oligo capping; fis (full insert sequence).		
SOURCE	Homo sapiens human lung cDNA to mRNA, clone_lib: LNG clone: LNG10890.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (sites) Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T., Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Tanigami,A., Fujiwara,T., Ono,T., Yamada,K., Fujii,Y., Ozaki,K., Hirao,M., Ohmori,Y., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.		
TITLE	NEDO human cDNA sequencing project		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 2129) Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.		
AUTHORS	Direct Submission		
TITLE	Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail: cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)		
JOURNAL	NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'-8,3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).		
FEATURES	Location/Qualifiers		
source	1. 2129 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="LNG10890" /tissue_type="human lung" /clone_lib="LNG" /note="cloning vector PME18SFL3"		
BASE COUNT	534 a 593 c 491 g 511 t		
ORIGIN			
Query Match	17.8%;	Score 633.8;	DB 9;
Best Local Similarity	60.1%;	Pred. No. 7.3e-138;	Length 2129;

Matches 1112; Conservative 0; Mismatches 722; Indels 15; Gaps 3;

Qy	1220	CTTGGCAATCCCTGTCCTCTCGAACAATTTTCAGATGGAAACCAAGAATGTAGACCATTGTC	1279
Db	33	CTGCGCAGCCCTGCCATATATGGTTGCTCTACTCCAATGG---CTCAGACTGTACCGCTGCC	89
Qy	1280	CAGCAGGAACGAGCCTGCACATTGGCTTTGAATAATAATGGTGAATGTCTCTCTCGCA	1339
Db	90	CTGAGGGAATGAACTGCTGTGGGATTTGAATACAAATGGTGGACACCGTGCCCAACA	149
Qy	1340	ACATGAAATCTTCCCTCTCAATGTGTGGAAATTCAAAGTCGGATGGAATGAATGGTTGGG	1399
Db	150	ACATGAAACGACCGTTCTCAGTGGGATCACTTCGAGCTCAAGGGCATGACAGGCTGGG	209
Qy	1400	AGGTGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTGACAAATGATTACCTGATCT	1459
Db	210	AGTGGCTGGTGATCACAATTTACACAGCTGCTGGAGCCTCAGACAAATGACTTCATGATC	269
Qy	1460	TAAACTTGCATATCCCAGGATTTAAACCAACCAACATCTATGACTGGAGCCAC---GGGTT	1516
Db	270	TCACTCTGGTTGGCCAGGATTTAGNCCCTCCGAGTCGGTGATGGCAGACACAGAGAATA	329
Qy	1517	CTGAAGTGAAGAAATACATTTGCTTTTGAGACCCCTCTGTTGAGCTGACTGATGTTTTGT	1576
Db	330	AAGAGGTGGCCAGAATACATTTGTCTTTGAGACCCCTCTGTTCTGTAACCTGTGAGCTCT	389
Qy	1577	ACTTCATGGTGGAATATTAATACAAAAGTACAAATGTCGTAGAAATCGTGGGTGGAAACCA	1636
Db	390	ACTTCATGGTGGTGGAATTTCTAGGCCAACACTCTCTGTGGAGACGTGGAAAGGTTCCA	449
Qy	1637	AAGAAAAAAGCTTACACCCATATCATCTTCAAGAATGCAACTTTTACATTTTACATGGG	1696
Db	450	AAGCAAAACAGTCTATACCTACATCATTTGAGGAGAACACTACCACGAGCTTCACTGGG	509
Qy	1697	CATTCCAGAGAACTATCAGGCTCAAGATAATAGACGGTTTCATCAATGACATGTTGAAGA	1756
Db	510	CCTTCCAGAGGACCACTTTTCTATGAGGCAAGCAGGAAGTACCAATAGACGTGGCCAAGA	569
Qy	1757	TTTATCTATCACAGCCACTAATGCAAGTTGATGGGGTGGCTCTCATGCGCTGCCTGTG	1816
Db	570	TCCTACTCCATCAATGTCCAAATGTTATGAATGGGTGGCTCTCTACTGCGTCCCTGTG	629
Qy	1817	CCCTCGGTTCTGAACAGTCGGGTTTCATGCTGTGCCCTGCCCTCCAGGCCACTACATTTG	1876
Db	630	CCCTAGAGCCCTGATGTGGGCTCCTCCTGCACCTCTTGTCTGCTGGTTATATATTG	689
Qy	1877	AGAAAGAAACCAACAGTGCAGGAATGTCCACCTGCACACCTACTCTGTCCTACATCAGG	1936
Db	690	ACCGAGATTCAGAACCTGCCACTCCTGCCCCCTTAACACAAATTCGTGAAAGCCACCAAG	749
Qy	1937	TCATATGGCAAGAGGCTGTATTCCTACCATGAAAAAGAAATCAGATTTTGCACATTG	1996
Db	750	CTTATGGTGTCAGGCTGTGTGCCCTGTGGTCTCAGGCAACCTCCAGCAAGATTCACCT	809
Qy	1997	CGGTTGCTTATAGTACTGCTTTTTTCTACCATGAAAAAGAAATCAGATTTTGCACATTG	2056
Db	810	CTCTGTGCTTACAAGATTTGCACCTTCTCAGCAACACTCCAGCAGGACTTTTCACTACA	869
Qy	2057	ACTTTAGCAACCTTCAGCAGTGTGGGCTCATTAATGAATGGCCCGCCAGGTTCACTTCAAG	2116
Db	870	ACTTCTCCGCTTTGGCAACACTGTCTACTCTTGTCTGGAGGCCAAGCTTCACTTCCAAG	929
Qy	2117	GAAACAATACTTCCATTTCTTCAATATCAGTTTATGTGGGCATGAGGGGAAGAAGATGG	2176
Db	930	GGCTGGAATACTTCCATCACTTTACCTTCAGTCTCTGTGGAAACCCAGGCTAGGAAAAATG	989
Qy	2177	CTCTCTCTACCAACATATAACAGACTTTTACAGTAAAAAGAAATAGTGGCAGGCTCAGATG	2236
Db	990	CTGTGTCACCAATATCTACTGACCTCCGGATTCCTTGAGGGTGAATCAGGG-----	1042
Qy	2237	ATTACAAAAATTTGTAGGGGCAATTTGATGCCACTCAACAATTTATCTCTTCTGAAAGTA	2296
Db	1043	--TTCTCCAAATCTATCACAGGCTACGTCTGCCAGGCACTCATCATCCCCCAGAGTGA	1100

Qy	2297	AGGTTTCCGAGACGCCCTTATCATCATCAAAATCCATTCATTCCTGGCAGATACATTCATAGGAG	2356
Db	1101	CAGGCTACAAGCGCGGGTTCCTCAGAGCCTGCAGCGTTCGCTCATCGACTTATTTGGGG	1160
Qy	2357	TCACAGTTTGAACACCAATGTGAATAATTAATAAAGAAGATATGTTCCAGTTCCAA	2416
Db	1161	TGACAACAGATATGACTCTGGATGGAAATCACTGCCACAGCTGAACATTTTTCACCTGGAGT	1220
Qy	2417	CAAGCCAATACCAGATCTGCATTTCTTTTATAAGTCTTCTACAGCAACAACATCTTTGTA	2476
Db	1221	CCTTGGGAATACCGGACGTGATCTCTCTTTTATAGTCCAATGATGTACCACAGCTCTGCA	1280
Qy	2477	TTAATGCGCGATCAACTGCTGTGAAAATGAGGTGTAACTTACTAAATCTGGAGCAGGAG	2536
Db	1281	GTCTTGGGAGATCAACCAACCATCCGCTCAGGTGCAGTCCACAGAAAACTCTCCCTGGAA	1340
Qy	2537	TGATTTTCACTCCCCAGCAAGTCCCCAGACAGTACCTCTGATGGGTGTACGTTCTATTTCC	2596
Db	1341	GTTTTGTCTGCTGCCAGGAACGTGCTCAGATGGGACCTGTGATGGCTGCAACTTTCACATTTCC	1400
Qy	2597	TGTGGGACAGTCTCAAGCTTTCGCCCTCTGTGTGAGGAGCATGACTTCCATGAGATTTGAGG	2656
Db	1401	TGTGGGAGCGCGGCTGCTTTGCCCGCTCTGCTCAGTGGCTGACTACCATGCTATCGTCA	1460
Qy	2657	GAGCCTGCAAGAGAGGATTTTCAGGAACCTTTGTATGTGTGGAAATGAACCTTAATTTGGTGA	2716
Db	1461	GCAGTGTGTGCTTGGGATCCAGAAGACTACTTACGTTGGCGAGAGACCCAAAGCTATGCT	1520
Qy	2717	TTAAGGAATTTCTTTGGCTCAGAAAAAGTTTGGCAACCTGTGAACGGTTTCACATTTTGGC	2776
Db	1521	CTGTGGCAATTTCTCTGCTGAGCAGAGAGTACCATCTGCAAAACCATAGATTTCTGCGC	1580
Qy	2777	TGAAGTGGGAGCCGGTGTGGAGCTTTTACTGCGGTTTGTGCTGGTGGCTGTGACCTGCT	2836
Db	1581	TCAAGTGGGCATCTCTCCAGGCACCTCTACTGCCATCCTGCTCACCGTCTTGGACCTGCT	1640
Qy	2837	ACTTCTGAAAAGAAATCAAAACTGGAAATACAAATATTCCAAGTTAGTAATGACGACTA	2896
Db	1641	ACTTTTGGAAAAGAAATCAAAACTAGAGTACAAAGTACTCCAAAGCTGGTGTGAATGTATA	1700
Qy	2897	ACTCAAAAGAGTGAACATCCCGGCTGCAGACAGTTGTGCTATCATGCAAGAGAGAAGATA	2956
Db	1701	CTCTCAAGGACTGTGACCTGGCAGCAGCTGACAGCTGCCCATCAATGGAGGCCGAGGATG	1760
Qy	2957	ATGAAGAGGAAGTTGTATATTTCCAATAAACAGTCACTACTAGGAAAACTCAAACTTTTGG	3016
Db	1761	TAGAGGACGACCTCATCTTACCAGCAAGAAGTCACCTCTTTTGGGAAGATCAAAATCATTTA	1820
Qy	3017	CAACCAAGAAAAGAAGACCAATTTTGAATCTGTGTTCAACTGAAACCTTC	3065
Db	1821	CTCTCAAGAGGACTCCTGATGGATTTGACTCAGTGGCGCTGAAGACATCT	1869

RESULT 13	
AB037745	
LOCUS	AB037745
DEFINITION	Homo sapiens mRNA for KIAA1324 protein, partial cds.
ACCESSION	AB037745
VERSION	AB037745.1
KEYWORDS	GI:7243028
SOURCE	Homo sapiens brain cDNA to mRNA, clone_llb:pBluescriptII SK plus clone:fh14139.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (sites)
AUTHORS	Nagase,T., Kikuno,R., Ishikawa,K.I., Hiroasawa,M. and Ohara,O.
TITLE	Prediction of the coding sequences of unidentified human genes. XVI. The complete sequences of 150 new cDNA clones from brain which code for large proteins in vitro
JOURNAL	DNA Res. 7 (1), 65-73 (2000)
MEDLINE	20181126

REFERENCE
AUTHORS
TITLE
JOURNAL

2 (bases 1 to 5567)
Ohta, O., Nagase, T. and Kikuno, R.
Direct Submission
Submitted (31-JAN-2000) Osamu Ohta, Kazusa DNA Research Institute,
Laboratory of DNA Technology, 1532-3 Yana, Kisarazu, Chiba
292-0812, Japan (E-mail: ocdnainfo@kazusa.or.jp,
URL: http://www.kazusa.or.jp/huge/, Tel: +81-438-52-3913,
Fax: +81-438-52-3914)

FEATURES

Location/Qualifiers
1..5567
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="fh14139"
/tissue_type="brain"
/clone_lib="pBluescriptII SK plus"
1..5567
/gene="KIAA1324"
c1..1744
/gene="KIAA1324"
/note="Start codon is not identified."
/codon_start=2
/product="KIAA1324 protein"
/protein_id="BA92562.1"
/db_xref="GI:7243029"
/translation="WNLTPTNMTTVLSGINFEYKGMTGWEVAGDHIYTAAGASND
FMILTVLPFRPQSQMDTENKEVARITFVETLCSVNCELYFMVGNSTRNTNPE
TWKSGKGSYTYIIEENTTSTFAFORTFHEASRKYNDVAKYISINVTNVMNGV
ASYRCPLAEADSGSSCTSCPAGYVIDRDSGTCHSCPNTILKAHOPXVOACVPCG
PGTKNNKIHSLCYNDCTFSRNPTRTFNFESALANTVYLAVLPGSPSTKGLYFHFHT
LSLCNGQRKMSVCTDNVIDRIPESGFSKSYTAIVCOAVIIPPEVYTGAGVSSQ
PVSLADRLJGVTDMLDGTSPAEFLHIESLGPDIPIFYRNDVYQSCSRSTTI
RVCSQPKTVPSGLLPGLGTCSDGDCGNFHEWASAAACPLCSVADYHAIYSSVAG
IOKTYVWREPKLCSGGLSLPEORVITCKTIDFWLKVGISAGTCTAILTLVLTCTYFWK
KNOKLEYKSKLVMMNATLKDCDLPADSDCAIMEGEDVEDDLIFTSKKSLFGKIKFSTS
KQAPVYISLSEDS"

BASE COUNT 1382 a 1404 c 1270 g 1511 t
ORIGIN

Query Match 16.7%; Score 594; DB 9; Length 5567;
Best Local Similarity 60.0%; Pred. No. 2e-128;
Matches 1031; Conservative 0; Mismatches 675; Indels 12; Gaps 2;
Qy 1317 ATGCTGAATGTCCTTCTGGCAACATGAAACATTCCTGCTTCAATGTTGGAAATTCAAA 1376
Db 1 ATGCTGAACACGCTGCCCAACACATGGAACGACCGTCTCAGTGGGATCAACTTCGA 60
Qy 1377 GTGCGATGGATGAATGGTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGCTGAGG 1436
Db 61 GTACAAGGGCATGACAGCTGGGAGGTGGCTGGTGTGATCATATTTACACAGCTGCTGGAGC 120
Qy 1437 TTCTGACAAATGATTACCTGATCTTAAACTTGATATCCAGGATTTAAACCAACCAATC 1496
Db 121 CTCAGACAAATGACTTCATGATCTCACTCTGGTGTGCCAGGATTTAGACCTCCGAGTC 180
Qy 1497 TATGACTGGAGCCAC---GGGTCTGAACTAGGAGAAATAACATTTGCTTTTGAGACCCCT 1553
Db 181 GGTGATGCGACACACAGAGAATAAAGAGTGGCGGAGAACATCATTTCTCTTGAGACCCCT 240
Qy 1554 CTGTTACAGTCGACTGCTGTTTGTACTTCATGGTGGATTAATAAGAAAAGATCAAAATGT 1613
Db 241 CTGTTCTGTGAACCTGTGAGCTCTACTTTTATTTTATTTTATTTTATTTTATTTTATTTT 300
Qy 1614 GGTAGAATCGTGGGTGGAACCAAGAAAACAAAGCTTTACACCCATATCATCTTCAAGAA 1673
Db 301 TGTGAGACGTGGAAAGGTTCCAAAGGCAACACAGCTCTCTATACCTATCATATGAGGAGAA 360
Qy 1674 TGCAACTTTTACATTTACATGGGATTTCCAGAGAACTAATCAGGGTCAAGATAATAGACG 1733
Db 361 CACTACCACGAGCTTCACTTGGGCTTCCAGAGGACCACCTTTTTCATGAGGCAACACGAA 420
Qy 1734 GTTCATCAATGACATGGTGAAGATTTATTTATTCACAGCCCACTAATCAGTTGATGGGT 1793
Db 421 GTACACCAATGACGTTGCCAAGATCTACTCCATCAATGTCACCAATGTTATGAATGCGGT 480

Qy 1794 GGGCTCTCATGCGCTGCGCTGCTGCCCTCGGTTCTGAACAGTCGGGTTCATCGTGTGCC 1853
Db 481 GGGCTCTCTACTGCGCTGCTGCCCTGTCCTAGAGCCTCTGATGTGGCTCTCCTGTCACCTC 540
Qy 1854 CTGCTCTCCAGGCACTACATNTGAGAAAGAAACACACAGTGCAGGAATGTCCACCTGA 1913
Db 541 TTCTCTGCTGCTGTTACTATATTTGACCGAGATTTCAGGAACCTCCCTGCGCCCTTAA 600
Qy 1914 CACCTACCTGTCTCATACATCAGCTCTATGCAAGAGGCTGTGATTCATCGCGGCTGG 1973
Db 601 CACAATCTGAAAGCCACAGCCTTATGTTGTCAGGCGCTGTGCTGCTGTGGTCCAGG 660
Qy 1974 GAGTAAAAACAATCAGGACCATTCGGTTCATATAGTACTGCTGCTTTTCTTACCATGAAA 2033
Db 661 GACCAAGAACAACAAGATCCACTCTCTGTCTACAAATGATTGACACCTTCTCACGCAACAC 720
Qy 2034 AGAAATCAGATTTTGCATATGACTTTTAGCAACCTCAGCAGTGTGGGCTCATTAATGAA 2093
Db 721 TCCAACACGAGCATTCAACTTACAACTTCTCCGCTTTTGGCAACACCGTCACTCTGCTGG 780
Qy 2094 TGGCCCCAGCTTCACCTCCAAAGGAACAATAATCTTCCATTTCTTCAATATCAGTTTATG 2153
Db 781 AGGCCAAGCTTCACCTCCAAAGGGTTGAAATACTTCCATCACTTTACCCCTCAGTCTCTG 840
Qy 2154 TGGGCATGAGGGGAAGAAGATGGCTCTCTGTACCAACAATATAACAGACTTTACAGTAAA 2213
Db 841 TGGAAACACAGGTAGGAAATGTCTGTGTGACCGACAATGTCTACGTCCCGGATTCC 900
Qy 2214 AGAAATAGTGGCAGGTGACATGATTACACAAATTTGGTAGGGGCAATTTGATGCCAGTC 2273
Db 901 TGAGGGTGTAGTCAGG-----TTCTCCAAATCTATCACAGCCTACGCTGCTGCCAGGC 951
Qy 2274 AACAAATATTCTCTCTGAAAGTAAAGGTTTCCGAGCAGCCTTATCATCAACAATCCATCAT 2333
Db 952 AGTCATCATCCCCCAGAGGTGACAGGCTACAAGGGCGGGGTTTCTTCACAGGCTGTGACG 1011
Qy 2334 TCTGGCAGATACATTTATAGGAGTACAGTTGAAACCAACATTTGAAAAATATTAATAAAA 2393
Db 1012 CCTTGTGATGACTTATTTGGGTGACACAGATATGACTCTGGATGGAATCACCCTCCC 1071
Qy 2394 AGAAGATATGTTCCAGTTCACAAAGCCAAATACACAGATGTGCAATTTCTTTTATAAGTC 2453
Db 1072 AGCTGAACTTTTCCACCTCGGAGTCTTGGGAATACCGGAGCTGATCTTCTTTTATAGGTC 1131
Qy 2454 TTCTACAGCAACACATCTTTGTTATTANTGCCGATCAACTGCTGTGAAATAGGTTGTA 2513
Db 1132 CAATGATGTACCCAGTCTCTGAGTTCCTGGGAGATCAACCAACCATCCGCGTCAGGTGACG 1191
Qy 2514 TCCTACTTAAATCTGAGCAGGAGTGATTTCAGTCCCCCAGCAAGTGCCAGCAGGATACCTG 2573
Db 1192 TCACAGAAACTGTCCCTGGAAGTTTGTCTGCCAGGAACGTGCTCAGATGGGACTG 1251
Qy 2574 TGATGGGTGACGTTCTTATTTCTGTGGAGAGTGTGAAAGCTTTGCCCTCTGTGTACGGA 2633
Db 1252 TGATGGGTGCAACTTCCACTTCTGTGGAGAGCGCGCTGCTTGGCCGCTCTGCTCAGT 1311
Qy 2634 GCATGACTTCCATCAGATGAGGGAGCCTGCAAGAGAGGATTTTCAGGAACCTTGTATGT 2693
Db 1312 GGCTGACTACATGCTATCTGACAGCTGTGTGGCTGGGATCCAGAAAGACTACTTACGT 1371
Qy 2694 GTGGAATGAACCTTAAATGGTGCATTAAGGAATTTCTTGTCTGAGAAAAAAGTTGGCAAC 2753
Db 1372 GTGCGAAGAACCCAGCTATGCTCTGGTGGCAATTTCTCTGCTGAGGAGAGATCACCAT 1431
Qy 2754 CTGTGAAACGTTGACTTTTGGCTGAGAGTGGGAGCGGTGTGGAGCTTTTACTGCCGT 2813
Db 1432 CTGCAAAACCATAGATTTCTGGGTGAAAGTGGGCACTCTCTGAGGACCTGTACTGCCAT 1491
Qy 2814 TTTGCTGGTGGCTGTGACCTGTCTACTTCTGAAAAAAGAAATCAAAAACTGGAATACAATA 2873
Db 1492 CCTGCTCACCCTCTTGTGCTGCTACTTCTTGGAAAAAGAAATCAAAAACTAGAGTACAAGTA 1551

QY 2874 TTCCAAGTTAGTATGACGACTAACTCAAAAGAGTGTGAATCCCGGCTGCAGACAGTTG 2933
 Db 1552 CTCCAAGCTGGTGTGATGATCTACTCTCAAGGACTGTGACCTGCCAGGCTGACAGCTG 1611
 QY 2934 TGCATATCATGAAGGAGAGATATGAAGAGGAGTTGTATATTCCTCAATAAACAGTCACT 2993.
 Db 1612 CGCCATCATGAAGCGAGGAGTGTAGAGGAGGAGCTCATCTTTACCAAGCAAGTCACT 1671
 QY 2994 ACTAGGAAACCTCAATCTTTGGCAACCAAGAAAAG 3031
 Db 1672 CTTTGGGAAGATCAATCAATTACCTCAAGCAGCCAG 1709

RESULT 14
 AC002081/c
 LOCUS AC002081 136150 bp DNA linear PRI 21-DEC-1999
 DEFINITION Homo sapiens BAC clone CTA-331C24 from 7q21, complete sequence.
 ACCESSION AC002081
 VERSION AC002081.1 GI:2078453
 KEYWORDS HTG.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Du.Z. and Maggi, L.
 TITLE The sequence of Homo sapiens BAC clone CTA-331C24
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 136150)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (12-MAY-1997) Department of Genetics, Washington
 University, 444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 REFERENCE 3 (bases 1 to 136150)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (24-SEP-1998) Department of Genetics, Washington
 University, 444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 REFERENCE 4 (bases 1 to 136150)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (21-DEC-1999) Department of Genetics, Washington
 University, 444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 COMMENT
 ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc>
 Contact: sapiens@watson.wustl.edu
 ----- Summary Statistics

 Center project name: H_RG331C24

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7> or <mailto:egreen@nhgri.nih.gov>

SOURCE INFORMATION:

Clone CTA-331C24 is from a release of the human BAC library CITB-HS-A. The library contains cloned DNA from human sperm. See: Shizuya et al., Proc.-Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (<http://www.resgen.com>).
 VECTOR: pBelOBAC11
 Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is CTA-369H2, 200 bp overlap.
 Actual start of this clone is at base position 1 of CTA-331C24;
 actual end is at 136150 of CTA-331C24.

FEATURES	Location/Qualifiers
source	1..136150
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="7"
	/map="7q21"
	/clone="CTA-331C24"
	/clone_lib="CITB-HS-A"
repeat_region	1270..1482
	/rpt_family="MIR"
repeat_region	4275..4325
	/rpt_family="L2"
repeat_region	4336..4419
	/rpt_family="MIR"
repeat_region	4446..4493
	/rpt_family="MIR"
repeat_region	4508..4567
	/rpt_family="L2"
repeat_region	7957..8098
	/rpt_family="L1"
repeat_region	9606..9740
	/rpt_family="MER1_type"
repeat_region	9781..9855
	/rpt_family="MER1_type"
repeat_region	10113..10230
	/rpt_family="L1"
repeat_region	10240..10396
	/rpt_family="L1"
repeat_region	10398..10679
	/rpt_family="Alu"
repeat_region	10712..11011
	/rpt_family="Alu"
repeat_region	11031..11230
	/rpt_family="L1"
repeat_region	11290..12047
	/rpt_family="MER2_type"
repeat_region	12037..12860
	/rpt_family="MER2_type"
repeat_region	13156..13350
	/rpt_family="Retroviral"
repeat_region	13356..13399
	/rpt_family="L1"
repeat_region	13600..14036
	/rpt_family="MaLR"
repeat_region	14108..14177
	/rpt_family="L1"
repeat_region	15128..15430
	/rpt_family="Alu"
repeat_region	16043..16278
	/rpt_family="L1"
repeat_region	18183..18331
	/rpt_family="L1"
repeat_region	18491..18791
	/rpt_family="Alu"
repeat_region	18807..19065
	/rpt_family="L1"
repeat_region	19612..19823
	/rpt_family="MIR"
gene	<21718..47235
CDS	/gene="WUGSC:H_RG331C24.1"
	join(<21718..22785,33250..33424,47162..47235)

```
/gene="WUGSC:H_RG331C24.1"
/note="match to Q14832 (PID:g1171564); H_RG331C24.1"
/codon_start=1
/product="metabotropic glutamate receptor type 3 (mGluR3)"
/protein_id="RAC60379.1"
/db_xref="GI:2078454"
/translation="APFNPKNDAISIVKFDFFGDMGRYVNFQNVGKYSYLKVGH
WAETLSLVNSIHMSRNVPSTQSDPCAPNEMKNMOPGDVCCWICIPCPEYELADE
FTCMDGSGQWPTADLTGCDLPEDYIRWEDAWAIGPVTIACLGFMCTCMVVTYFIKH
NNTPLVKAGRELVCYLLFGLSVCMPTFFFIKAPSPVICALRLGLGSSFAICYSAI
LTKTNCIARIQDKVNGAORKEFISPSOVFICLGLILLVQIMVSWLLLEAPGTRY
TLAKRETVILKCNVKSLSLITLYIVILVILCTVYAFKTRKCPENENEARFIFGTFM
YTTCILWIAFPIFYTSDSYRQVTTMCIIISVLSGFWVLGLGFAPKVIHILFOPQKN
VWTRHLNRFVSQSGTGTYSQSSASTYVPTVCNGRELDSTSSL"
repeat_region
22949..23079
/rpt_family="L2"
/rpt_family="L2"
23798..24042
/rpt_family="L1"
/rpt_family="L1"
24919..25091
/rpt_family="MERL_type"
25657..25937
/gene="WUGSC:H_RG331C24.1"
/db_xref="GI:1113596"
26579..26649
/rpt_family="L2"
26771..26919
/rpt_family="MERL_type"
27205..27582
/rpt_family="L2"
27771..28069
/rpt_family="Alu"
30158..30625
/rpt_family="L2"
30909..30980
/rpt_family="L2"
33325..33424
/gene="WUGSC:H_RG331C24.1"
/note="match to EST AA670430 (NID:g2631929) ad20g04.sl"
complement(33345..33424)
/note="match to EST (NID:g1568674)"
33557..33660
/rpt_family="L2"
34927..35033
/rpt_family="MIR"
35868..35981
/rpt_family="L2"
36503..36746
/rpt_family="MaLR"
37496..37634
/rpt_family="MERL_type"
39042..39351
/rpt_family="Alu"
39357..39479
/rpt_family="L1"
40234..40486
/rpt_family="L2"
40835..41135
/rpt_family="Alu"
41202..41565
/rpt_family="MaLR"
41844..42141
/rpt_family="Alu"
44395..45111
/rpt_family="L2"
46236..46352
/rpt_family="MaLR"
46911..47118
/rpt_family="L1"
complement(47141..47518)
```

Query Match 14.6%; Score 518.4; DB 9; Length 136150;

Best Local Similarity 99.8%; Pred. No. 2.4e-110;

Matches 519; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 3023 AGGAAAAGAGACCATTTTGAATCTGTTCAACTGAAACCTCAAGATCCCCAAATATAT 3082
|||||
Db 63412 AGGAAAAGAGACCATTTTGAATCTGTTCAACTGAAACCTCAAGATCCCCAAATATAT 63353
QY 3083 GAAGACAGCTCTGTAGCTTGAGACTAATGAACAAGAAACCTGCTCTAGTTTACAG 3142
|||||
Db 63352 GAAGACAGCTCTGTAGCTTGAGACTAATGAACAAGAAACCTGCTCTAGTTTACAG 63293
QY 3143 GACCATTATTTAGGCTGCTCTCACTACCTGTCACATTTGGTGTATCTCAGAGAGAGGCC 3202
|||||
Db 63292 GACCATTATTTAGGCTGCTCTCACTACCTGTCACATTTGGTGTATCTCAGAGAGAGGCC 63233
QY 3203 ATGCCGCTGAAAAGGAAGAGAGATTGAAACATTTGATTTGCCCTTATCACATGGTCAAGTAC 3262
|||||
Db 63232 ATGCCGCTGAAAAGGAAGAGAGATTGAAACATTTGATTTGCCCTTATCACATGGTCAAGTAC 63173
QY 3263 CTTGCCAATAAAGAAAGCAATGATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAG 3322
|||||
Db 63172 CTTGCCAATAAAGAAAGCAATGATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAG 63113
QY 3323 GAAGAGATTATCTGTATATACACATACTGAAACCAAGTTTAAAGCCCAACTGCACCT 3382
|||||
Db 63112 GAAGAGATTATCTGTATATACACATACTGAAACCAAGTTTAAAGCCCAACTGCACCT 63053
QY 3383 GCTGATGTCATGCATATTAATGGGTAACTTTTATCTTTATGATGTCTACATAACAA 3442
|||||
Db 63052 GCTGATGTCATGCATATTAATGGGTAACTTTTATCTTTATGATGTCTACATAACAA 62993
QY 3443 GTGTGATTTGGAGGCACATGTGAGCATATGATATGATATGATATGATTTTCTTTT 3502
|||||
Db 62992 GTGTGATTTGGAGGCACATGTGAGCATATGATATGATATGATATGATTTTCTTTT 62933
QY 3503 GTTTATATTTGGGAAAATTAATAATTTTAAAGTAAA 3542
|||||
Db 62932 GTTTATATTTGGGAAAATTAATAATTTTAAAGTATA 62893

RESULT 15
AX336856/c
LOCUS AX336856 576 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 7365 from Patent WO0194629.
ACCESSION AX336856
VERSION AX336856.1 GI:18127575
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
Horriagan,S., Soppet,D.R. and Weaver,Z.
TITLE Cancer gene determination and therapeutic screening using signature
JOURNAL Patent: WO 0194629-A 7365 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
Source
1..576
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 171 a 112 c 108 g 182 t 3 others
ORIGIN

Query Match 14.2%; Score 504; DB 6; Length 576;
Best Local Similarity 94.4%; Pred. No. 1.7e-107;
Matches 543; Conservative 0; Mismatches 28; Indels 4; Gaps 2;

QY 2970 TGTATATTCGAATAAAGTCACCTACTAGAAAACCTCAATCTTTGGCAACCAAGGAAA 3029
|||||
Db 575 TGTATATTCGAATAAAGTCACCTACTAGAAAACCTCAATCTTTGGCAACCAAGGAAA 516
|||||
QY 3030 AGAAGA---CCATTTTGAATCTGTT--CAACTGAAACCTCAAGATCCCCAAATATATGAA 3085
|||||
```

Db	515	AAAGGAGACCCATTTTGAATCTGTTCCAACTGAAACCCCTCAAGATVCCCAAAATATATGAA	456
Qy	3086	GAGACAGTGTGTAGCTTTGAGACTAATGAACAAAGAAACCTGCTAGTATTTTACAGGAC	3145
Db	455	GAGACAGTCTGTAGCCCTTTGAGACTAATGANCAGGAAACCTGCTCTAGTATTTTACAGGAC	396
Qy	3146	CATATTTTAGGGTCTGCTCATACCTGTCACATTTGGTGATCTCACAGAGGAGGGCCATG	3205
Db	395	CATATTTTAGGGTCTGCTCTCATACCTGTCACATTTGGTGATCTCACAGAGGAGGGCCATG	336
Qy	3206	CCGCTGAAAAGGGAAGGAGATTGAAACATTTTGATTCGCTTATACATGGTCAAGTACCTT	3265
Db	335	CCGCTGAAAAGGGAAGGAGATTGAAACATTTTGATTCGCTTATACATGGTCAAGTACCTT	276
Qy	3266	GCCAAATAAGGGAAGCAAAATGATTTGGTCTCAACTGAAGTGAAGCTCAACTCAGGAA	3325
Db	275	GCCAAATAAGGGAAGCAAAATGATTTGGTCTCAACTGAAGTGAAGCTCAACTCAGGAA	216
Qy	3326	GAGATTTATCTGTATATACACATAACTGAAAACCAAGTTTAAGCCCAACCAATGCACCTGCT	3385
Db	215	GAGATTTATCTGTATATACACATAACTGNAACCAAGTTTAAGCCCAACCAATGCACCTGCT	156
Qy	3386	GATGCATGCCATATAATTAATGGTAACTTTTATCTTTATGATGCTACATAACAAGTG	3445
Db	155	GATGCATGCCATATAATTAATGGTAACTTTTATCTTTATGATGCTACATAACAAGTG	96
Qy	3446	TGATTTGGAAGGCACATGTGACCATATGCATTTATGATCCAAATTTATCTTTTCTTTGTT	3505
Db	95	TGATTTGGAAGGCACATGTGACCATATGCATTTATGATCCAAATTTATCTTTTCTTTGTT	36
Qy	3506	TATATTTTGGGGAANAATTAATAATTTTTTTTTTAAGGTA	3540
Db	35	TATATTTTGGGGAANAATTAATAATTTTTTTTTTTAAAGTA	1

Search completed: May 12, 2003, 05:35:10
Job time : 9096.17 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 13:40:59 ; Search time 18.8729 Seconds
(without alignments)
4695.671 Million cell updates/sec

Title: US-10-073-333A-2
Perfect score: 963
Sequence: 1 MLFRARGVRGWRGWRPAEA.....TCYFWKKKQKKKKTILNLFN 963

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 349150 seqs, 92025710 residues

Word size : 0

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_AA.*

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	963	100.0	963	9	US-10-176-921-38
2	953	99.0	1027	9	US-10-140-164-2
3	271	28.1	464	9	US-10-002-050-20
4	271	28.1	464	9	US-10-002-304-20
5	271	28.1	464	12	US-10-003-152-20
6	232	24.1	411	9	US-10-002-050-10
7	232	24.1	411	9	US-10-002-304-10
8	232	24.1	411	12	US-10-003-152-10
9	81	8.4	81	10	US-09-864-761-39769
10	74	7.7	74	9	US-10-140-164-32
11	74	7.7	74	9	US-10-140-164-61
12	64	6.6	64	10	US-09-864-761-47095
13	60	6.2	60	10	US-09-864-761-39057
14	15	1.6	1013	9	US-10-028-072-38
15	15	1.6	1013	9	US-10-121-049-38
16	15	1.6	1013	9	US-10-123-904-38
17	15	1.6	1013	9	US-10-140-470-38
18	15	1.6	1013	9	US-10-175-746-38
19	15	1.6	1013	9	US-10-176-918-38

20	15	1.6	1013	9	US-10-176-921-38	Sequence 38, Appl
21	15	1.6	1013	9	US-10-137-865-38	Sequence 38, Appl
22	15	1.6	1013	9	US-10-140-474-38	Sequence 38, Appl
23	15	1.6	1013	9	US-10-142-431-38	Sequence 38, Appl
24	15	1.6	1013	9	US-10-143-114-38	Sequence 38, Appl
25	15	1.6	1013	9	US-10-140-002-38	Sequence 38, Appl
26	15	1.6	1013	9	US-10-142-419-38	Sequence 38, Appl
27	15	1.6	1013	9	US-10-123-262-38	Sequence 38, Appl
28	15	1.6	1013	9	US-10-142-423-38	Sequence 38, Appl
29	15	1.6	1013	9	US-10-121-050-38	Sequence 38, Appl
30	15	1.6	1013	9	US-10-141-755-38	Sequence 38, Appl
31	15	1.6	1013	9	US-10-143-032-38	Sequence 38, Appl
32	15	1.6	1013	9	US-10-123-108-38	Sequence 38, Appl
33	15	1.6	1013	9	US-10-123-236-38	Sequence 38, Appl
34	15	1.6	1013	9	US-10-123-261-38	Sequence 38, Appl
35	15	1.6	1013	9	US-10-140-921-38	Sequence 38, Appl
36	15	1.6	1013	9	US-10-140-928-38	Sequence 38, Appl
37	15	1.6	1013	9	US-10-121-045-38	Sequence 38, Appl
38	15	1.6	1013	9	US-10-123-292-38	Sequence 38, Appl
39	15	1.6	1013	9	US-10-123-903-38	Sequence 38, Appl
40	15	1.6	1013	9	US-10-124-819-38	Sequence 38, Appl
41	15	1.6	1013	9	US-10-124-822-38	Sequence 38, Appl
42	15	1.6	1013	9	US-10-140-925-38	Sequence 38, Appl
43	15	1.6	1013	9	US-10-160-498-38	Sequence 38, Appl
44	15	1.6	1013	9	US-10-121-041-38	Sequence 38, Appl
45	15	1.6	1013	9	US-10-121-043-38	Sequence 38, Appl

ALIGNMENTS

RESULT 1
US-10-140-164-2
; Sequence 2, Application US/10140164
; Publication No. US20030072736A1
; GENERAL INFORMATION:
; APPLICANT: Baker et al.
; TITLE OF INVENTION: Human Tumor Necrosis Factor Receptor TR16
; FILE REFERENCE: PF514C1
; CURRENT APPLICATION NUMBER: US/10/140,164
; CURRENT FILING DATE: 2002-05-08
; PRIOR APPLICATION NUMBER: 09/637,856
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/148,348
; PRIOR FILING DATE: 1999-08-12
; PRIOR APPLICATION NUMBER: 60/148,683
; PRIOR FILING DATE: 1999-08-13
; PRIOR APPLICATION NUMBER: 60/148,870
; PRIOR FILING DATE: 1999-08-13
; PRIOR APPLICATION NUMBER: 60/148,758
; PRIOR FILING DATE: 1999-08-16
; PRIOR APPLICATION NUMBER: 60/149,181
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: 60/149,453
; PRIOR FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: 60/149,498
; PRIOR FILING DATE: 1999-08-19
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 963
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-140-164-2

Query Match 100.0%; Score 963; DB 9; Length 963;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 963; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFRARGVRGWRGWRPAEAPRRGRSPWPSPAWICWALAGCOAAGDLPSSSSRPLPP 60
|||||
DB 1 MLFRARGVRGWRGWRPAEAPRRGRSPWPSPAWICWALAGCOAAGDLPSSSSRPLPP 60
|||||

```
Qy 61 COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120
Db 61 COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120
Qy 121 SKCGGTSLGSGIKFDEWDELPAFNSIATFMDTVVGPDSRDPGCCNNSWIPRGNYIE 180
Db 121 SKCGGTSLGSGIKFDEWDELPAFNSIATFMDTVVGPDSRDPGCCNNSWIPRGNYIE 180
Qy 181 SNRDDCTVSLIYAVHLKSGYVFEFYQYVDNNIFFEFFFQNDQCQEMDTTDDKWVKLTDN 240
Db 181 SNRDDCTVSLIYAVHLKSGYVFEFYQYVDNNIFFEFFFQNDQCQEMDTTDDKWVKLTDN 240
Qy 241 GEMGSHSVMLKSGTNILYWRRTGILMGSKAVKPVLVKNITIEGVAYTSECPCPKCTFSN 300
Db 241 GEMGSHSVMLKSGTNILYWRRTGILMGSKAVKPVLVKNITIEGVAYTSECPCPKCTFSN 300
Qy 301 KPGSFNCQVCPRNTYSEKAKECIRCKDDSQFSGSECTERPPCTTKDYFQIHTPCDEEG 360
Db 301 KPGSFNCQVCPRNTYSEKAKECIRCKDDSQFSGSECTERPPCTTKDYFQIHTPCDEEG 360
Qy 361 KTOIMYKWIETPKICREDLTDALRLPPSGEKKDCPCNPGFYNNGSSSCHPCPPGTFSDGT 420
Db 361 KTOIMYKWIETPKICREDLTDALRLPPSGEKKDCPCNPGFYNNGSSSCHPCPPGTFSDGT 420
Qy 421 KECRCPAGTEPALGFEYKWNVLPCNMKTSCFNVGNKSCDGMNGWEVAGDHIQSGAGS 480
Db 421 KECRCPAGTEPALGFEYKWNVLPCNMKTSCFNVGNKSCDGMNGWEVAGDHIQSGAGS 480
Qy 481 DNDYLILNLHIPGFKPPTSMTCATGSELGRITFVFETLCSADCVLYFMVDINRKSTNVVE 540
Db 481 DNDYLILNLHIPGFKPPTSMTCATGSELGRITFVFETLCSADCVLYFMVDINRKSTNVVE 540
Qy 541 SWGGTKEKQAYTHIIFKNATFTTFAFORTNOQDNRREFINDMWKIYSTATNAVGVAS 600
Db 541 SWGGTKEKQAYTHIIFKNATFTTFAFORTNOQDNRREFINDMWKIYSTATNAVGVAS 600
Qy 601 SCRACALGSEQSGSCVPCPPGHYIEKETNOCKECPPTDYLIIHQYVGEACIPCGPGSK 660
Db 601 SCRACALGSEQSGSCVPCPPGHYIEKETNOCKECPPTDYLIIHQYVGEACIPCGPGSK 660
Qy 661 NNQDHSVCISDCFFHYEKEKNQILHVDFSNLSSVGLMNGPSTSKTKYFHFENISLGH 720
Db 661 NNQDHSVCISDCFFHYEKEKNQILHVDFSNLSSVGLMNGPSTSKTKYFHFENISLGH 720
Qy 721 EGKKNALCTNNITDFTVKEIVAGSDDYTNLGAFCVQSTIIPSESKGFRALSSQSIIILA 780
Db 721 EGKKNALCTNNITDFTVKEIVAGSDDYTNLGAFCVQSTIIPSESKGFRALSSQSIIILA 780
Qy 781 DTFIGVTVETTLKINIKEDMFPVPTSQIPDVHFFYKSTATTSCINGRSTAVKRCNPT 840
Db 781 DTFIGVTVETTLKINIKEDMFPVPTSQIPDVHFFYKSTATTSCINGRSTAVKRCNPT 840
Qy 841 KSGAGVISVPSKCPAGTCDCGCTFYFLWESAECPLCTEHDHFEIEGACKRGQETLYVNN 900
Db 841 KSGAGVISVPSKCPAGTCDCGCTFYFLWESAECPLCTEHDHFEIEGACKRGQETLYVNN 900
Qy 901 EPRKWCIGISLPEKKLATCETVDFWLKVGAGFAFTAVLLVALTYCFWKNNQKKKTIILN 960
Db 901 EPRKWCIGISLPEKKLATCETVDFWLKVGAGFAFTAVLLVALTYCFWKNNQKKKTIILN 960
Qy 961 LFN 963
Db 961 LFN 963
```

RESULT 2

US-10-140-164-4
; Sequence 4, Application US/10140164
; Publication NO. US20030072736A1
; GENERAL INFORMATION:
; APPLICANT: Baker et al.
; TITLE OF INVENTION: Human Tumor Necrosis Factor Receptor TR16
; FILE REFERENCE: PF514C1

ut

```
; CURRENT APPLICATION NUMBER: US/10/140,164  
; CURRENT FILING DATE: 2002-05-08  
; PRIOR APPLICATION NUMBER: 09/637,856  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/148,348  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR APPLICATION NUMBER: 60/148,683  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: 60/148,870  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: 60/148,758  
; PRIOR FILING DATE: 1999-08-16  
; PRIOR APPLICATION NUMBER: 60/149,181  
; PRIOR FILING DATE: 1999-08-17  
; PRIOR APPLICATION NUMBER: 60/149,453  
; PRIOR FILING DATE: 1999-08-18  
; PRIOR APPLICATION NUMBER: 60/149,498  
; PRIOR FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 76  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 1027  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-140-164-4
```

Query Match 99.0%; Score 953; DB 9; Length 1027;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 953; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 MLFRARGPVRGWGRPAEAPRRGSRPPWPAWICWALAGCOAAWAGDLPSSSSRPLPP 60
Db 1 MLFRARGPVRGWGRPAEAPRRGSRPPWPAWICWALAGCOAAWAGDLPSSSSRPLPP 60
Qy 61 COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120
Db 61 COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120
Qy 121 SKCGGTSLGSGIKFDEWDELPAFNSIATFMDTVVGPDSRDPGCCNNSWIPRGNYIE 180
Db 121 SKCGGTSLGSGIKFDEWDELPAFNSIATFMDTVVGPDSRDPGCCNNSWIPRGNYIE 180
Qy 181 SNRDDCTVSLIYAVHLKSGYVFEFYQYVDNNIFFEFFFQNDQCQEMDTTDDKWVKLTDN 240
Db 181 SNRDDCTVSLIYAVHLKSGYVFEFYQYVDNNIFFEFFFQNDQCQEMDTTDDKWVKLTDN 240
Qy 241 GEMGSHSVMLKSGTNILYWRRTGILMGSKAVKPVLVKNITIEGVAYTSECPCPKCTFSN 300
Db 241 GEMGSHSVMLKSGTNILYWRRTGILMGSKAVKPVLVKNITIEGVAYTSECPCPKCTFSN 300
Qy 301 KPGSFNCQVCPRNTYSEKAKECIRCKDDSQFSGSECTERPPCTTKDYFQIHTPCDEEG 360
Db 301 KPGSFNCQVCPRNTYSEKAKECIRCKDDSQFSGSECTERPPCTTKDYFQIHTPCDEEG 360
Qy 361 KTOIMYKWIETPKICREDLTDALRLPPSGEKKDCPCNPGFYNNGSSSCHPCPPGTFSDGT 420
Db 361 KTOIMYKWIETPKICREDLTDALRLPPSGEKKDCPCNPGFYNNGSSSCHPCPPGTFSDGT 420
Qy 421 KECRCPAGTEPALGFEYKWNVLPCNMKTSCFNVGNKSCDGMNGWEVAGDHIQSGAGS 480
Db 421 KECRCPAGTEPALGFEYKWNVLPCNMKTSCFNVGNKSCDGMNGWEVAGDHIQSGAGS 480
Qy 481 DNDYLILNLHIPGFKPPTSMTCATGSELGRITFVFETLCSADCVLYFMVDINRKSTNVVE 540
Db 481 DNDYLILNLHIPGFKPPTSMTCATGSELGRITFVFETLCSADCVLYFMVDINRKSTNVVE 540
Qy 541 SWGGTKEKQAYTHIIFKNATFTTFAFORTNOQDNRREFINDMWKIYSTATNAVGVAS 600
Db 541 SWGGTKEKQAYTHIIFKNATFTTFAFORTNOQDNRREFINDMWKIYSTATNAVGVAS 600
Qy 601 SCRACALGSEQSGSCVPCPPGHYIEKETNOCKECPPTDYLIIHQYVGEACIPCGPGSK 660
Db 601 SCRACALGSEQSGSCVPCPPGHYIEKETNOCKECPPTDYLIIHQYVGEACIPCGPGSK 660
```

QY 661 NNQDSVCSYDCFFYHEKENQILHYDFSNLSSVGLMNGPSFTSKGTYKYPHFHFNISLCGH 720
DB 661 NNQDSVCSYDCFFYHEKENQILHYDFSNLSSVGLMNGPSFTSKGTYKYPHFHFNISLCGH 720
QY 721 EGKKNALCTNNITDFTVKEIVAGSDYTNLVGAFVCSQSTIIPSESKGFRALSSQSILIA 780
DB 721 EGKKNALCTNNITDFTVKEIVAGSDYTNLVGAFVCSQSTIIPSESKGFRALSSQSILIA 780
QY 781 DFIGVTVETLLKINIKEDMFVPYTSQIPDVHFFYKSTATTSCINGRSTAVMRNCPT 840
DB 781 DFIGVTVETLLKINIKEDMFVPYTSQIPDVHFFYKSTATTSCINGRSTAVMRNCPT 840
QY 841 KSGAGVISVPSKCPAGTCDGCTFFYLWESAECPLCTEHDHFEIEGACKRGFQETLYWYN 900
DB 841 KSGAGVISVPSKCPAGTCDGCTFFYLWESAECPLCTEHDHFEIEGACKRGFQETLYWYN 900
QY 901 EPKWCIGISLPEKKLATCETVDFWLKVGAGVGAFTAVLLVALTCYFWKKNOK 953
DB 901 EPKWCIGISLPEKKLATCETVDFWLKVGAGVGAFTAVLLVALTCYFWKKNOK 953

RESULT 3
US-10-002-050-20
; Sequence 20, Application US/10002050
; Publication No. US20030032095A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corine
; APPLICANT: Yang, Meijia
; APPLICANT: Boldog, Ferenc
; APPLICANT: Herrmann, John
; TITLE OF INVENTION: Polynucleotides and polypeptides encoded thereby
; FILE REFERENCE: 15966-554 Cura-54 CON-S14
; CURRENT APPLICATION NUMBER: US/10/002,050
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 09/604,286
; PRIOR FILING DATE: 2000-06-22
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-002-050-20

Query Match 28.1%; Score 271; DB 9; Length 464;
Best Local Similarity 99.8%; Pred. No. 1.6e-257;
Matches 451; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 115 MKNQVCSKCGEGTYSLSGKIKFDEWDELPAFNSNIATFMDTVVGPDSRDPDGCNNSWIP 174
DB 1 MKNQVCSKCGEGTYSLSGKIKFDEWDELPAFNSNIATFMDTVVGPDSRDPDGCNNSWIP 60
QY 175 RGNYESNRDDCTVSLIYAVHLKSGYVFFEQYVDNNIIEFFFIQNDQCEMDTTTDKW 234
DB 61 RGNYESNRDDCTVSLIYAVHLKSGYVFFEQYVDNNIIEFFFIQNDQCEMDTTTDKW 120
QY 235 VKLTDNGEWSHVSMLKSGTNILYWRITGILMGSKAVKPVLVKNITIEGVAYTSECFPC 294
DB 121 VKLTDNGEWSHVSMLKSGTNILYWRITGILMGSKAVKPVLVKNITIEGVAYTSECFPC 180
QY 295 PGTFNSKPGSFNCQVCPNTYSEKGAKEICRCKDDSQFS--GSSECTERPPCTTKDYFQI 352
DB 181 PGTFNSKPGSFNCQVCPNTYSEKGAKEICRCKDDSQFSSEGSSECTERPPCTTKDYFQI 240
QY 353 HTPCDEBEGTKQIMYKWIPEKICREDLTDAILRPPSGEKKDCPCNPGFYNNSSSCHPCP 412
DB 241 HTPCDEBEGTKQIMYKWIPEKICREDLTDAILRPPSGEKKDCPCNPGFYNNSSSCHPCP 300

QY 115 MKNQVCSKCGEGTYSLSGKIKFDEWDELPAFNSNIATFMDTVVGPDSRDPDGCNNSWIP 174
DB 1 MKNQVCSKCGEGTYSLSGKIKFDEWDELPAFNSNIATFMDTVVGPDSRDPDGCNNSWIP 60
QY 175 RGNYESNRDDCTVSLIYAVHLKSGYVFFEQYVDNNIIEFFFIQNDQCEMDTTTDKW 234
DB 61 RGNYESNRDDCTVSLIYAVHLKSGYVFFEQYVDNNIIEFFFIQNDQCEMDTTTDKW 120
QY 235 VKLTDNGEWSHVSMLKSGTNILYWRITGILMGSKAVKPVLVKNITIEGVAYTSECFPC 294
DB 121 VKLTDNGEWSHVSMLKSGTNILYWRITGILMGSKAVKPVLVKNITIEGVAYTSECFPC 180
QY 295 PGTFNSKPGSFNCQVCPNTYSEKGAKEICRCKDDSQFS--GSSECTERPPCTTKDYFQI 352
DB 181 PGTFNSKPGSFNCQVCPNTYSEKGAKEICRCKDDSQFSSEGSSECTERPPCTTKDYFQI 240
QY 353 HTPCDEBEGTKQIMYKWIPEKICREDLTDAILRPPSGEKKDCPCNPGFYNNSSSCHPCP 412
DB 241 HTPCDEBEGTKQIMYKWIPEKICREDLTDAILRPPSGEKKDCPCNPGFYNNSSSCHPCP 300

QY 413 PGTFSDGTKECRPCPAGTEPALGFEYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDH 472
DB 301 PGTFSDGTKECRPCPAGTEPALGFEYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDH 360
QY 473 IQSGAGSDNDYLILNLHIPGFKPPTSMTGATGSELGRITTFVETLCSADCVLYFMDIN 532
DB 361 IQSGAGSDNDYLILNLHIPGFKPPTSMTGATGSELGRITTFVETLCSADCVLYFMDIN 420
QY 533 RKSTNVVESGGTKEKQAYTHIIFKNATFTFTW 565
DB 421 RKSTNVVESGGTKEKQAYTHIIFKNATFTFTW 453

RESULT 4
US-10-002-304-20
; Sequence 20, Application US/10002304
; Publication No. US20030036185A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corine
; APPLICANT: Yang, Meijia
; APPLICANT: Boldog, Ferenc
; APPLICANT: Herrmann, John
; TITLE OF INVENTION: Polynucleotides and polypeptides encoded thereby
; FILE REFERENCE: 15966-554 Cura-54 CON-S8
; CURRENT APPLICATION NUMBER: US/10/002,304
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 09/604,286
; PRIOR FILING DATE: 2000-06-22
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-002-304-20

Query Match 28.1%; Score 271; DB 9; Length 464;
Best Local Similarity 99.6%; Pred. No. 1.6e-257;
Matches 451; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 115 MKNQVCSKCGEGTYSLSGKIKFDEWDELPAFNSNIATFMDTVVGPDSRDPDGCNNSWIP 174
DB 1 MKNQVCSKCGEGTYSLSGKIKFDEWDELPAFNSNIATFMDTVVGPDSRDPDGCNNSWIP 60
QY 175 RGNYESNRDDCTVSLIYAVHLKSGYVFFEQYVDNNIIEFFFIQNDQCEMDTTTDKW 234
DB 61 RGNYESNRDDCTVSLIYAVHLKSGYVFFEQYVDNNIIEFFFIQNDQCEMDTTTDKW 120
QY 235 VKLTDNGEWSHVSMLKSGTNILYWRITGILMGSKAVKPVLVKNITIEGVAYTSECFPC 294
DB 121 VKLTDNGEWSHVSMLKSGTNILYWRITGILMGSKAVKPVLVKNITIEGVAYTSECFPC 180
QY 295 PGTFNSKPGSFNCQVCPNTYSEKGAKEICRCKDDSQFS--GSSECTERPPCTTKDYFQI 352
DB 181 PGTFNSKPGSFNCQVCPNTYSEKGAKEICRCKDDSQFSSEGSSECTERPPCTTKDYFQI 240
QY 353 HTPCDEBEGTKQIMYKWIPEKICREDLTDAILRPPSGEKKDCPCNPGFYNNSSSCHPCP 412
DB 241 HTPCDEBEGTKQIMYKWIPEKICREDLTDAILRPPSGEKKDCPCNPGFYNNSSSCHPCP 300
QY 413 PGTFSDGTKECRPCPAGTEPALGFEYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDH 472
DB 301 PGTFSDGTKECRPCPAGTEPALGFEYKWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDH 360
QY 473 IQSGAGSDNDYLILNLHIPGFKPPTSMTGATGSELGRITTFVETLCSADCVLYFMDIN 532
DB 361 IQSGAGSDNDYLILNLHIPGFKPPTSMTGATGSELGRITTFVETLCSADCVLYFMDIN 420
QY 533 RKSTNVVESGGTKEKQAYTHIIFKNATFTFTW 565

Db 421 RKSTNVESWGTEKQAYTHIFKNATFTFTW 453
 |||

RESULT 5

US-10-003-152-20

; Sequence 20, Application US/100003152

; Patent No. US20020151494A1

; GENERAL INFORMATION:

; APPLICANT: Shimkets, Richard

; APPLICANT: Fernandes, Elma

; APPLICANT: Vernet, Corine

; APPLICANT: Yang, Meijia

; APPLICANT: Boldog, Ferenc

; APPLICANT: Herrmann, John

; TITLE OF INVENTION: No. US20020151494A1el Amino Acid Sequences for Human Semaphorin-1

; FILE REFERENCE: 15966-554 Cura-54 CON-S12

; CURRENT APPLICATION NUMBER: US/10/003,152

; CURRENT FILING DATE: 2001-11-02

; PRIOR APPLICATION NUMBER: 09/604,286

; PRIOR FILING DATE: 2000-06-22

; PRIOR APPLICATION NUMBER: 60/140,584

; PRIOR FILING DATE: 1999-06-23

; NUMBER OF SEQ ID NOS: 49

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 20

; LENGTH: 464

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-003-152-20

Query Match 28.1%; Score 271; DB 12; Length 464;

Best Local Similarity 99.6%; Pred. No. 1.6e-257;

Matches 431; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 115 MKQVSKCGEGTYSLSGSGIKFDEWDELPAFGSNIAETMDTVVGPDSRDPDCNNSWIP 174
 |||

Db 1 MKQVSKCGEGTYSLSGSGIKFDEWDELPAFGSNIAETMDTVVGPDSRDPDCNNSWIP 60
 |||

QY 175 RGNVIESNRDDCTVSLIYAVHLKKSQYVFEEYQYVDNNIFFEFFIIONDQCQEMDITTDKW 234
 |||

Db 61 RGNVIESNRDDCTVSLIYAVHLKKSQYVFEEYQYVDNNIFFEFFIIONDQCQEMDITTDKW 120
 |||

QY 235 VKLTDNGEMGSHVMLKSGTNILYRTTILMGSKAVKPVLVKNITIEGVATYSECFPC 294
 |||

Db 121 VKLTDNGEMGSHVMLKSGTNILYRTTILMGSKAVKPVLVKNITIEGVATYSECFPC 180
 |||

QY 295 PGTFNKPQSFNCQVCPRTYSEKGAKECIRCKDDSQFS--GSSECTERPPCTTKDYFQI 352
 |||

Db 181 PGTFNKPQSFNCQVCPRTYSEKGAKECIRCKDDSQFSSEGSSECTERPPCTTKDYFQI 240
 |||

QY 353 HTPCDEGKQIMYKIEWPKICREDLTDAILPPSGEKKDCPCPNPGFYNGSSSCHPCP 412
 |||

Db 241 HTPCDEGKQIMYKIEWPKICREDLTDAILPPSGEKKDCPCPNPGFYNGSSSCHPCP 300
 |||

QY 413 PGTFSDGTEKRCPCPAGTEPALGFYKWMNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDH 472
 |||

Db 301 PGTFSDGTEKRCPCPAGTEPALGFYKWMNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDH 360
 |||

QY 473 IQSGAGGSDNDYLIINLHPIPGKPPPTSMTGATGSELGRITFVFETLCSADCVLYFMVDIN 532
 |||

Db 361 IQSGAGGSDNDYLIINLHPIPGKPPPTSMTGATGSELGRITFVFETLCSADCVLYFMVDIN 420
 |||

QY 533 RKSTNVESWGTEKQAYTHIFKNATFTFTW 565
 |||

Db 421 RKSTNVESWGTEKQAYTHIFKNATFTFTW 453
 |||

RESULT 6

US-10-002-050-10

; Sequence 10, Application US/10002050

; Publication No. US20030032095A1

; GENERAL INFORMATION:

; APPLICANT: Shimkets, Richard
 ; APPLICANT: Fernandes, Elma
 ; APPLICANT: Vernet, Corine
 ; APPLICANT: Yang, Meijia
 ; APPLICANT: Boldog, Ferenc
 ; APPLICANT: Herrmann, John

; TITLE OF INVENTION: No. US20030032095A1el Nucleic Acid Sequences Encoding Human Se

; FILE REFERENCE: 15966-554 Cura-54 CON-S14

; CURRENT APPLICATION NUMBER: US/10/002,050

; CURRENT FILING DATE: 2001-11-02

; PRIOR APPLICATION NUMBER: 09/604,286

; PRIOR FILING DATE: 2000-06-22

; PRIOR APPLICATION NUMBER: 60/140,584

; PRIOR FILING DATE: 1999-06-23

; NUMBER OF SEQ ID NOS: 49

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 10

; LENGTH: 411

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-002-050-10

Query Match 24.1%; Score 232; DB 9; Length 411;

Best Local Similarity 100.0%; Pred. No. 2.9e-219;

Matches 232; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 GSSECTERPPCTTKDYFQIHPTCDEEGKQIMYKIEWPKICREDLTDAILRPPSGEKKDC 393
 |||

Db 169 GSSECTERPPCTTKDYFQIHPTCDEEGKQIMYKIEWPKICREDLTDAILRPPSGEKKDC 228
 |||

QY 394 PCPNPGFYNGSSSCHPCPPGTFSDGTEKRCPCPAGTEPALGFYKWMNVLPGNMKTSCF 453
 |||

Db 229 PCPNPGFYNGSSSCHPCPPGTFSDGTEKRCPCPAGTEPALGFYKWMNVLPGNMKTSCF 288
 |||

QY 454 NVGNSKCDGMNGWEVAGDHIOGAGGSDNDYLIINLHPIPGKPPPTSMTGATGSELGRITF 513
 |||

Db 289 NVGNSKCDGMNGWEVAGDHIOGAGGSDNDYLIINLHPIPGKPPPTSMTGATGSELGRITF 348
 |||

QY 514 VFETLCSADCVLYFMVDINRKSTNVESWGTEKQAYTHIFKNATFTFTW 565
 |||

Db 349 VFETLCSADCVLYFMVDINRKSTNVESWGTEKQAYTHIFKNATFTFTW 400
 |||

RESULT 7

US-10-002-304-10

; Sequence 10, Application US/10002304

; Publication No. US20030036185A1

; GENERAL INFORMATION:

; APPLICANT: Shimkets, Richard

; APPLICANT: Fernandes, Elma

; APPLICANT: Vernet, Corine

; APPLICANT: Yang, Meijia

; APPLICANT: Boldog, Ferenc

; APPLICANT: Herrmann, John

; TITLE OF INVENTION: Polynucleotides and polypeptides encoded thereby

; FILE REFERENCE: 15966-554 Cura-54 CON-S8

; CURRENT APPLICATION NUMBER: US/10/002,304

; CURRENT FILING DATE: 2001-11-02

; PRIOR APPLICATION NUMBER: 09/604,286

; PRIOR FILING DATE: 2000-06-22

; PRIOR APPLICATION NUMBER: 60/140,584

; PRIOR FILING DATE: 1999-06-23

; NUMBER OF SEQ ID NOS: 49

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 10

; LENGTH: 411

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-002-304-10

Query Match

Best Local Similarity 100.0%; Pred. No. 2.9e-219;

Matches 232; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 GSSECTERPPCTTKDYFOIHTPCDEGKTOIMYKWIPIKICREDLTDALRLPPSGEKKDC 393
Db 169 GSSECTERPPCTTKDYFOIHTPCDEGKTOIMYKWIPIKICREDLTDALRLPPSGEKKDC 228
QY 394 PCNPGFYNNSSSCHPCPPGTFSDGTKECRPCPAGTEPALGFEYKWNVLPGNMKTSF 453
Db 229 PCNPGFYNNSSSCHPCPPGTFSDGTKECRPCPAGTEPALGFEYKWNVLPGNMKTSF 288
QY 454 NVGNSKCDGMNGWEVAGDHQSGAGSDNDYLILNLHIPGFKPPTSMGTGATGSELGRITF 513
Db 289 NVGNSKCDGMNGWEVAGDHQSGAGSDNDYLILNLHIPGFKPPTSMGTGATGSELGRITF 348
QY 514 VFETLCSADCVLFYFMVDINRKSTNNVESWGGTKEKOAYTHIIFKNATFTFTW 565
Db 349 VFETLCSADCVLFYFMVDINRKSTNNVESWGGTKEKOAYTHIIFKNATFTFTW 400

RESULT 8
US-10-003-152-10
; Sequence 10, Application US/10003152
; Patent No. US20020151494A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corine
; APPLICANT: Yang, Meijia
; APPLICANT: Boldog, Ferenc
; APPLICANT: Herrmann, John
; TITLE OF INVENTION: No. US20020151494A1el Amino Acid Sequences for Human Semaphorin-1
; FILE REFERENCE: 15966-554 Cura-54 CON-S12
; CURRENT APPLICATION NUMBER: US/10/003,152
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 09/604,286
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/140,584
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 411
; TYPE: PRT
; ORGANISM: Homo sapiens.
US-10-003-152-10

Query Match 24.1%; Score 232; DB 12; Length 411;
Best Local Similarity 100.0%; Pred. No. 2.9e-219;
Matches 232; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 334 GSSECTERPPCTTKDYFOIHTPCDEGKTOIMYKWIPIKICREDLTDALRLPPSGEKKDC 393
Db 169 GSSECTERPPCTTKDYFOIHTPCDEGKTOIMYKWIPIKICREDLTDALRLPPSGEKKDC 228
QY 394 PCNPGFYNNSSSCHPCPPGTFSDGTKECRPCPAGTEPALGFEYKWNVLPGNMKTSF 453
Db 229 PCNPGFYNNSSSCHPCPPGTFSDGTKECRPCPAGTEPALGFEYKWNVLPGNMKTSF 288
QY 454 NVGNSKCDGMNGWEVAGDHQSGAGSDNDYLILNLHIPGFKPPTSMGTGATGSELGRITF 513
Db 289 NVGNSKCDGMNGWEVAGDHQSGAGSDNDYLILNLHIPGFKPPTSMGTGATGSELGRITF 348
QY 514 VFETLCSADCVLFYFMVDINRKSTNNVESWGGTKEKOAYTHIIFKNATFTFTW 565
Db 349 VFETLCSADCVLFYFMVDINRKSTNNVESWGGTKEKOAYTHIIFKNATFTFTW 400

RESULT 9
US-09-864-761-39769
; Sequence 39769, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Weisheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecmica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 39769
; LENGTH: 81
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC002081.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN LONG, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
; OTHER INFORMATION: EST-HUMAN HIT: AW954806.1, EVALUE 5.00e-45
; OTHER INFORMATION: SWISSPROT HIT: P01267, EVALUE 3.00e-03
US-09-864-761-39769

Query Match 8.4%; Score 81; DB 10; Length 81;
Best Local Similarity 100.0%; Pred. No. 1.5e-71;
Matches 81; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 578 RFINDMVKIYSITATNAVDGVASSCRACALGSEQSSGVPCPPGHIYEKTNQCKECP 637
Db 1 RFINDMVKIYSITATNAVDGVASSCRACALGSEQSSGVPCPPGHIYEKTNQCKECP 60
QY 638 DTYLSTHQVYKGEACIPCGPG 658
Db 61 DTYLSTHQVYKGEACIPCGPG 81

RESULT 10
US-10-140-164-32
; Sequence 32, Application US/10140164
; Publication No. US20030072736A1
; GENERAL INFORMATION:
; APPLICANT: Baker et al.
; TITLE OF INVENTION: Human Tumor Necrosis Factor Receptor Trl6
; FILE REFERENCE: PF514C1
; CURRENT APPLICATION NUMBER: US/10/140,164
; CURRENT FILING DATE: 2002-05-08
; PRIOR APPLICATION NUMBER: 09/637,856
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/148,348
; PRIOR FILING DATE: 1999-08-12
; PRIOR APPLICATION NUMBER: 60/148,683
; PRIOR FILING DATE: 1999-08-13
; PRIOR APPLICATION NUMBER: 60/148,870
; PRIOR FILING DATE: 1999-08-13
; PRIOR APPLICATION NUMBER: 60/148,758
; PRIOR FILING DATE: 1999-08-16
; PRIOR APPLICATION NUMBER: 60/149,181
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: 60/149,453
; PRIOR FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: 60/149,498
; PRIOR FILING DATE: 1999-08-19
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 32
; LENGTH: 74
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-140-164-32

Query Match
Best Local Similarity 100.0%; Pred. No. 1e-64;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 338 CTERPCTTKDYFQIHTPCDEEGKTQIMYKWIPEKICREDLTDALRLPPSGEKKDCPPCN 397
Db 1 CTERPCTTKDYFQIHTPCDEEGKTQIMYKWIPEKICREDLTDALRLPPSGEKKDCPPCN 60

QY 398 PGFYNNNGSSSCHPC 411
Db 61 PGFYNNNGSSSCHPC 74

RESULT 11
US-10-140-164-61
; Sequence 61, Application US/10140164
; Publication No. US20030072736A1
; GENERAL INFORMATION:
; APPLICANT: Baker et al.
; TITLE OF INVENTION: Human Tumor Necrosis Factor Receptor Trl6
; FILE REFERENCE: PF514C1
; CURRENT APPLICATION NUMBER: US/10/140,164
; CURRENT FILING DATE: 2002-05-08
; PRIOR APPLICATION NUMBER: 09/637,856
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/148,348
; PRIOR FILING DATE: 1999-08-12
; PRIOR APPLICATION NUMBER: 60/148,683
; PRIOR FILING DATE: 1999-08-13
; PRIOR APPLICATION NUMBER: 60/148,870
; PRIOR FILING DATE: 1999-08-13
; PRIOR APPLICATION NUMBER: 60/148,758
; PRIOR FILING DATE: 1999-08-16
; PRIOR APPLICATION NUMBER: 60/149,181
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: 60/149,453
; PRIOR FILING DATE: 1999-08-18

; PRIOR APPLICATION NUMBER: 60/149,498
; PRIOR FILING DATE: 1999-08-19
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 61
; LENGTH: 74
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-140-164-61.

Query Match
Best Local Similarity 100.0%; Pred. No. 1e-64;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 338 CTERPCTTKDYFQIHTPCDEEGKTQIMYKWIPEKICREDLTDALRLPPSGEKKDCPPCN 397
Db 1 CTERPCTTKDYFQIHTPCDEEGKTQIMYKWIPEKICREDLTDALRLPPSGEKKDCPPCN 60

QY 398 PGFYNNNGSSSCHPC 411
Db 61 PGFYNNNGSSSCHPC 74

RESULT 12
US-09-864-761-47095
; Sequence 47095, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29

; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 47095
; LENGTH: 64
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC002081.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.55
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.52
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.49
; OTHER INFORMATION: SWISSPROT HIT: Q00019, EVALUE 8.90e-01
US-09-864-761-47095

Query Match 6.6%; Score 64; DB 10; Length 64;
Best Local Similarity 100.0%; Pred. No. 6.1e-55;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 105 FSCASGEYLEMKNQVCSKCGCTYSLGSGIKFDEWDELPAGFSNIATFMDTVVGPDSRSP 164
Db 1 FSCASGEYLEMKNQVCSKCGCTYSLGSGIKFDEWDELPAGFSNIATFMDTVVGPDSRSP 60

Qy 165 DGCN 168
Db 61 DGCN 64

RESULT 13
US-09-864-761-39057
; Sequence 39057, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Acomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21

; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 39057
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC002081.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6
; OTHER INFORMATION: EST HUMAN HIT: AW966212.1, EVALUE 2.00e-10
; OTHER INFORMATION: SWISSPROT HIT: P21849, EVALUE 5.00e-03
US-09-864-761-39057

Query Match 6.2%; Score 60; DB 10; Length 60;
Best Local Similarity 100.0%; Pred. No. 4.9e-51;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 362 TQIMYKWIPEKICREDLTDAIRLPPSGEKKDCPCNCPGYNNGSSSCHPCPPGTGSDGK 421
Db 1 TQIMYKWIPEKICREDLTDAIRLPPSGEKKDCPCNCPGYNNGSSSCHPCPPGTGSDGK 60

RESULT 14
US-10-028-072-38
; Sequence 38, Application US/10028072
; Publication No. US20030004311A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang
; TITLE OF INVENTION:
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/028,072
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263

1 PRIOR FILING DATE: 1997-09-18
 2 PRIOR APPLICATION NUMBER: 60/059352
 3 PRIOR FILING DATE: 1997-09-19
 4 PRIOR APPLICATION NUMBER: 60/059588
 5 PRIOR FILING DATE: 1997-09-19
 6 PRIOR APPLICATION NUMBER: 60/059836
 7 PRIOR FILING DATE: 1997-09-24
 8 PRIOR APPLICATION NUMBER: 60/062250
 9 PRIOR FILING DATE: 1997-10-17
 10 PRIOR APPLICATION NUMBER: 60/062285
 11 PRIOR FILING DATE: 1997-10-17
 12 PRIOR APPLICATION NUMBER: 60/062287
 13 PRIOR FILING DATE: 1997-10-17
 14 PRIOR APPLICATION NUMBER: 60/062814
 15 PRIOR FILING DATE: 1997-10-24
 16 PRIOR APPLICATION NUMBER: 60/062816
 17 PRIOR FILING DATE: 1997-10-24
 18 PRIOR APPLICATION NUMBER: 60/063045
 19 PRIOR FILING DATE: 1997-10-24
 20 PRIOR APPLICATION NUMBER: 60/063082
 21 PRIOR FILING DATE: 1997-10-31
 22 PRIOR APPLICATION NUMBER: 60/063127
 23 PRIOR FILING DATE: 1997-10-24
 24 PRIOR APPLICATION NUMBER: 60/063327
 25 PRIOR FILING DATE: 1997-10-27
 26 PRIOR APPLICATION NUMBER: 60/063329
 27 PRIOR FILING DATE: 1997-10-27
 28 PRIOR APPLICATION NUMBER: 60/063550
 29 PRIOR FILING DATE: 1997-10-28
 30 PRIOR APPLICATION NUMBER: 60/063561
 31 PRIOR FILING DATE: 1997-10-28
 32 PRIOR APPLICATION NUMBER: 60/063704
 33 PRIOR FILING DATE: 1997-10-29
 34 PRIOR APPLICATION NUMBER: 60/063733
 35 PRIOR FILING DATE: 1997-10-29
 36 PRIOR APPLICATION NUMBER: 60/063735
 37 PRIOR FILING DATE: 1997-10-29
 38 PRIOR APPLICATION NUMBER: 60/063738
 39 PRIOR FILING DATE: 1997-10-29
 40 PRIOR APPLICATION NUMBER: 60/063755
 41 PRIOR FILING DATE: 1997-10-17
 42 PRIOR APPLICATION NUMBER: 60/064248
 43 PRIOR FILING DATE: 1997-11-03
 44 PRIOR APPLICATION NUMBER: 60/064809
 45 PRIOR FILING DATE: 1997-11-07
 46 PRIOR APPLICATION NUMBER: 60/065186
 47 PRIOR FILING DATE: 1997-11-12
 48 PRIOR APPLICATION NUMBER: 60/065846
 49 PRIOR FILING DATE: 1997-11-17
 50 PRIOR APPLICATION NUMBER: 60/066364
 51 PRIOR FILING DATE: 1997-11-21
 52 PRIOR APPLICATION NUMBER: 60/066453
 53 PRIOR FILING DATE: 1997-11-24
 54 PRIOR APPLICATION NUMBER: 60/066511
 55 PRIOR FILING DATE: 1997-11-24
 56 PRIOR APPLICATION NUMBER: 60/066770
 57 PRIOR FILING DATE: 1997-11-24
 58 PRIOR APPLICATION NUMBER: 60/069212
 59 PRIOR FILING DATE: 1997-12-11
 60 PRIOR APPLICATION NUMBER: 60/069278
 61 PRIOR FILING DATE: 1997-12-11
 62 PRIOR APPLICATION NUMBER: 60/069334
 63 PRIOR FILING DATE: 1997-12-11
 64 PRIOR APPLICATION NUMBER: 60/069694
 65 PRIOR FILING DATE: 1997-12-16
 66 PRIOR APPLICATION NUMBER: 60/072320
 67 PRIOR FILING DATE: 1998-01-23
 68 PRIOR APPLICATION NUMBER: 60/073612
 69 PRIOR FILING DATE: 1998-02-04
 70 PRIOR APPLICATION NUMBER: 60/074086
 71 PRIOR FILING DATE: 1998-02-09
 72 PRIOR APPLICATION NUMBER: 60/074092
 73 PRIOR FILING DATE: 1998-02-09

Qy	283	GVAYTSECFFCKPGT	297
Db	268	GVAYTSECFFCKPGT	282

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 13:23:58 ; Search time 13.0658 Seconds
(without alignments)
3056.959 Million cell updates/sec

Title: US-10-073-333A-2

Perfect score: 963

Sequence: 1 MLFRAGPVRGRGCRPAEA.....TCYFWKKKKKKKTIILNLFN 963

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	DEF6_HUMAN	8	0.8	100	1	Q01524 homo sapien
2	TNR4_RAT	8	0.8	271	1	P15725 rattus norv
3	TNR4_MOUSE	8	0.8	272	1	P47741 mus musculus
4	VBL1_TGMV	8	0.8	293	1	P03566 tomato gold
5	TRGB_HUMAN	8	0.8	300	1	Q95407 homo sapien
6	DCP2_PEA	8	0.8	405	1	P51851 pisum sativ
7	DCP1_TOBAC	8	0.8	418	1	P51845 nicotiana t
8	TRLT_HUMAN	8	0.8	430	1	Q96924 homo sapien
9	TRLT_WACEFA	8	0.8	430	1	Q9n092 macaca fasc
10	DCP3_ORYSA	8	0.8	585	1	P51849 oryza sativ
11	DCP2_ORYSA	8	0.8	603	1	P51848 oryza sativ
12	KRP2_RAT	8	0.8	671	1	O62909 rattus norv
13	MCAK_CRIGR	8	0.8	718	1	P70096 cricetus
14	TAT_SIVCZ	7	0.7	100	1	P17285 chimpanzee
15	TK24_GUITH	7	0.7	101	1	O46905 guillardia
16	CRE2_BACHD	7	0.7	134	1	Q9k819 bacillus ha
17	NEUV_RANES	7	0.7	141	1	P11858 rana escul
18	MMS2_MYCTU	7	0.7	145	1	Q11170 mycobacteri
19	NEUV_BURJA	7	0.7	159	1	P08163 bufo japoni
20	YCBQ_ECOLI	7	0.7	179	1	P73855 escherichia
21	YD85_METJA	7	0.7	179	1	Q58780 methanococ
22	RS8A_SCHPO	7	0.7	200	1	O14049 schizosacch
23	RS8B_SCHPO	7	0.7	200	1	Q9p7b2 schizosacch
24	Y296_METJA	7	0.7	203	1	O57744 methanococ
25	Y930_HAEIN	7	0.7	206	1	P44077 naemophilus
26	RASM_MOUSE	7	0.7	208	1	O08989 mus musculus
27	RASM_RAT	7	0.7	208	1	P97538 rattus norv
28	YM16_YEAST	7	0.7	208	1	P40206 saccharomyc
29	YQED_BACSU	7	0.7	208	1	P54449 bacillus su
30	SP07_YEAST	7	0.7	259	1	P18410 saccharomyc
31	PYRF_CANAL	7	0.7	270	1	P13649 candida alb
32	PYRF_CANDU	7	0.7	270	1	Q9c150 candida dub
33	ATND_XENLA	7	0.7	277	1	P21188 xenopus lae

34 7 0.7 283 1 TRI4_HUMAN Q92956 homo sapien
35 7 0.7 320 1 DNC_HUMAN Q9hc21 homo sapien
36 7 0.7 335 1 FLIG_THEMA Q9wv63 thermotoga
37 7 0.7 362 1 YD33_MYCLE P53425 mycobacteri
38 7 0.7 365 1 WN14_HUMAN O14904 homo sapien
39 7 0.7 418 1 YE87_SCHPO O14303 schizosacch
40 7 0.7 420 1 YCEL_CAEEL Q94175 caenorhabdi
41 7 0.7 437 1 YQEZ_BACSU P54465 bacillus su
42 7 0.7 459 1 G33_RAT P50432 rattus norv
43 7 0.7 463 1 YAD4_YEAST P28003 saccharomyc
44 7 0.7 529 1 YPCL_CAEEL O11178 caenorhabdi
45 7 0.7 548 1 TRM1_SCHPO Q9p804 schizosacch

ALIGNMENTS

RESULT 1

ID DEF6_HUMAN STANDARD; PRT; 100 AA.
AC Q01524;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Defensin 6 precursor (Defensin, alpha 6).
GN DEFA6 OR DEF6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Intestine;
RX MEDLINE=93114459; PubMed=8417977;
RA Jones D.E., Bevins C.L.;
RT "Defensin-6 mRNA in human Paneth cells: implications for
antimicrobial peptides in host defense of the human bowel.";
RL FEBS Lett. 315:187-192(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96223969; PubMed=8626737;
RA Mallow E.B., Harris A., Salzman N., Russell J.P.,
RA Deberardinis R.J., Ruchelli E., Bevins C.L.;
RT "Human enteric defensins. Gene structure and developmental
expression.";
RL J. Biol. Chem. 271:4038-4045(1996).
CC -!- FUNCTION: HAS ANTIMICROBIAL ACTIVITY.
CC -!- TISSUE SPECIFICITY: PANETH CELLS OF THE SMALL INTESTINE.
CC -!- SIMILARITY: BELONGS TO THE CORTICOSTATIN/DEFENSIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

CC EMBL; M98331; AAB59357.1;
CC EMBL; U33317; AAC50382.1; ALT_SEQ.
CC PIR; S27016; S27016.
CC Genew; HGNC:2765; DEFA6.
CC MIM; 600471;
CC InterPro; IPR002366; Defensin_alpha.
CC InterPro; IPR001271; Defensin_mammal.
CC Pfam; PF00323; defensins; 1.
CC Pfam; PF00879; Defensin_propep; 1.
CC SMART; SM00048; DEFSN; 1.
CC PROSITE; PS00269; DEFENSIN; 1.
CC Defensin; Antibiatic; Fungicide; Signal.
KW SIGNAL 1 19 POTENTIAL.
FT PROPEP 20 ?65 POTENTIAL.
FT CHAIN ?66 100 DEFENSIN 6.

```
FT DISULFID 72 99 BY SIMILARITY.
FT DISULFID 74 98 BY SIMILARITY.
FT DISULFID 78 98 BY SIMILARITY.
SQ SEQUENCE 100 AA; 10975 MW; EDF77E033DDCE2D5 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 100;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 936 TAVLLVAL 943
DB 8 TAVLLVAL 15

RESULT 2
TNR4_RAT
ID TNR4_RAT STANDARD; PRT; 271 AA.
AC P15725;
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tumor necrosis factor receptor superfamily member 4 precursor (OX40L
DE receptor) (OX40 antigen) (MRC OX40).
GN TNFRSF4 OR TNGPIL OR OX40.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=T-cell;
RX MEDLINE=90214614; PubMed=2157591;
RA Mallett S., Fossum S., Barclay A.N.;
RT "Characterization of the MRC OX40 antigen of activated CD4 positive T
RT lymphocytes -- a molecule related to nerve growth factor receptor.";
RL EMBO J. 9:1063-1068(1990).
CC -!- FUNCTION: Receptor for TNFSF4/OX40L/GP34.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: ACTIVATED T-CELLS.
CC -!- SIMILARITY: CONTAINS 4 TNFR-CYS REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X17037; CAA34897.1; -.
CC PIR; S08036; S08036.
CC PIR; S12783; S12783.
CC HSSP; O14763; 1D4V.
CC InterPro: IPR001368; TNFR_c6.
CC Pfam; PF00020; TNFR_c6; 3.
CC ProDom; PD000771; TNFR_c6; 1.
CC SMART; SM00208; TNFR; 3.
CC PROSITE; PS00652; TNFR_NGFR_1; 2.
CC PROSITE; PS50050; TNFR_NGFR_2; 2.
CC Receptor; Antigen; Transmembrane; Glycoprotein; Repeat; Signal:
FT SIGNAL 1 19
FT CHAIN 20 271
FT DOMAIN 20 210
FT TRANSMEM 21 235
FT DOMAIN 236 271
FT REPEAT 25 60
FT REPEAT 61 102
FT REPEAT 103 123
FT REPEAT 124 164
FT REPEAT 165 207
FT DISULFID 26 37
FT DISULFID 38 51
FT DISULFID 41 59
```

```
FT DISULFID 62 76 BY SIMILARITY.
FT DISULFID 79 94 BY SIMILARITY.
FT DISULFID 82 102 BY SIMILARITY.
FT DISULFID 104 122 BY SIMILARITY.
FT DISULFID 125 138 BY SIMILARITY.
FT DISULFID 144 163 BY SIMILARITY.
FT CARBOHYD 143 143 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 271 AA; 29895 MW; C06465136B16E821 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 271;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCPGGH 623
DB 122 CVPCPGGH 129

RESULT 3
TNR4_MOUSE
ID TNR4_MOUSE STANDARD; PRT; 272 AA.
AC P47741;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tumor necrosis factor receptor superfamily member 4 precursor (OX40L
DE receptor) (OX40 antigen).
GN TNFRSF4 OR TNGP1 OR OX40.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RX MEDLINE=94044750; PubMed=8228223;
RA Calderhead D.M., Buhlmann J.E., van den Eertwegh A.J.,
RA Barclay A.N.;
RT "Cloning of mouse OX40: a T cell activation marker that may mediate
RT T-B cell interactions.";
RL J. Immunol. 151:5261-5271(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95255413; PubMed=7737295;
RA Birkeland M.L., Copeland N.G., Gilbert D.J., Jenkins N.A.,
RA Barclay A.N.;
RT "Gene structure and chromosomal localization of the mouse homologue
RT of rat OX40 protein.";
RL Eur. J. Immunol. 25:926-930(1995).
CC -!- FUNCTION: Receptor for TNFSF4/OX40L/GP34.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: CONTAINS 4 TNFR-CYS REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z21674; CAA79772.1; -.
CC EMBL; X85214; CAA59476.1; -.
CC HSSP; O14763; 1D0G.
CC MGD; MGI:104512; Tnfrsf4.
CC InterPro: IPR001368; TNFR_c6.
CC Pfam; PF00020; TNFR_c6; 3.
CC ProDom; PD000771; TNFR_c6; 1.
CC SMART; SM00208; TNFR; 3.
CC PROSITE; PS00652; TNFR_NGFR_1; 2.
CC PROSITE; PS50050; TNFR_NGFR_2; 2.
CC Receptor; Antigen; Transmembrane; Glycoprotein; Repeat; Signal:
FT SIGNAL 1 19
FT CHAIN 20 272
```

FT CHAIN 20 272 TUMOR NECROSIS FACTOR RECEPTOR
FT DOMAIN 20 211 SUPERFAMILY MEMBER 4.
FT TRANSMEM 212 236 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 237 272 POTENTIAL.
FT REPEAT 26 61 CYTOPLASMIC (POTENTIAL).
FT REPEAT 62 103 TNFR-CYS 1.
FT REPEAT 104 124 TNFR-CYS 2.
FT REPEAT 125 165 TNFR-CYS 3 (INCOMPLETE).
FT DISULFID 27 38 BY SIMILARITY.
FT DISULFID 39 52 BY SIMILARITY.
FT DISULFID 42 60 BY SIMILARITY.
FT DISULFID 63 77 BY SIMILARITY.
FT DISULFID 80 95 BY SIMILARITY.
FT DISULFID 93 103 BY SIMILARITY.
FT DISULFID 105 123 BY SIMILARITY.
FT DISULFID 126 139 BY SIMILARITY.
FT DISULFID 145 164 BY SIMILARITY.
FT CARBOHYD 144 144 N-LINKED (GLCNAC...) (POTENTIAL).
FT CONFLICT 15 15 A -> G (IN REF. 2).
SQ SEQUENCE 272 AA: 30153 MW: 06E7BB4156FD08E CRC64;

Query Match 0.8%; Score 8; DB 1; Length 272;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCPPGH 623
DB 123 CVPCPPGH 130

RESULT 4.

ID VBL1_TGMV STANDARD; PRT; 293 AA.
AC P03566;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE BL1 protein.
OS BCL1.
GN Tomato golden mosaic virus (TGMV).
OC Viruses; ssDNA viruses; Geminiviridae; Begomovirus.
OX NCBI_TaxID=10831;
RN [1]
RP SEQUENCE FROM N.A.
RA von Arnim A.G., Stanley J.;
RL Submitted (XXX-1991) to the EMBL/GenBank/DBSJ databases.
RN [2]

PRELIMINARY SEQUENCE FROM N.A.
RA Hamilton W.D.O., Stein V.E., Coutts R.H.A., Buck K.W.;
RT "Complete nucleotide sequence of the infectious cloned DNA components of tomato golden mosaic virus: potential coding regions and regulatory sequences."
RL EMBO J. 3:2197-2205(1984).
CC -!- SIMILARITY: BELONGS TO GEMINIVIRUSES BL1 PROTEIN FAMILY.
CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 173 ONWARD DUE TO A FRAMESHIFT ERROR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----

DR EMBL; M73794; AAA46584.1; -
DR EMBL; K02030; -; NOT_ANNOTATED_CDS.
DR PIR; A04169; QOCVLG.
DR InterPro; IPR000211; Gemini.BL.
DR Pfam; PF00845; Gemini.BL.1; v -> A (IN REF. 2).

FT CONFLICT 11 11
SQ SEQUENCE 293 AA: 32933 MW: E024E3F285C89012 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 293;
Best Local Similarity 100.0%; Pred. No. 5.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 177 NYIESNRD 184
DB 13 NYIESNRD 20

RESULT 5

ID TR6B_HUMAN STANDARD; PRT; 300 AA.
AC O95407;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tumor necrosis factor receptor superfamily member 6B precursor (Decoy receptor for Fas ligand) (Decoy receptor 3) (Dcr3) (M68).
GN TNFRSF6B OR DCR3 OR TR6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal lung;
RX MEDLINE=99087326; PubMed=9872321;
RA Pitti R.M., Marsters S.A., Lawrence D.A., Roy M., Kischkel F.C., Dowd P., Huang A., Donahue C.J., Sherwood S.W., Baldwin D.T., Godowski P.J., Wood W.I., Gurney A.L., Hillan K.J., Cohen R.L., Goddard A.D., Botstein D., Ashkenazi A.;
RT "Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer";
RL Nature 396:699-703(1998).
RN [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 30-35.
RC TISSUE=Prostate;
RX MEDLINE=99253915; PubMed=10318773;
RA Yu K.-Y., Kwon B., Ni J., Zhai Y., Ebner R., Kwon B.S.;
RT "A newly identified member of tumor necrosis factor receptor superfamily (TR6) suppresses LIGHT-mediated apoptosis";
RL J. Biol. Chem. 274:13733-13736(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=20122600; PubMed=10655513;
RA Bai C., Connolly B., Metzker M.L., Hilliard C.A., Liu X., Sandig V., Soderman A., Galloway S.M., Liu Q., Austin C.P., Caskey C.T.;
RT "Overexpression of M68/Dcr3 in human gastrointestinal tract tumors independent of gene amplification and its location in a four-gene cluster";
RL Proc. Natl. Acad. Sci. U.S.A. 97:1230-1235(2000).
RN [4]
RP SEQUENCE FROM N.A.
RA Matthews L.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBSJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX Strausberg R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBSJ databases.

-!- FUNCTION: Decoy receptor for the cytotoxic ligands TNFSF14/LIGHT and TNFSF6/FasL. Protects against apoptosis.
-!- SUBCELLULAR LOCATION: Secreted.
-!- TISSUE SPECIFICITY: Detected in fetal lung, brain and liver. Detected in adult stomach, spinal cord, lymph node, trachea, spleen, colon and lung. Highly expressed in several primary tumors from colon, stomach, rectum, esophagus and in SW480 colon carcinoma cells.
-!- SIMILARITY: CONTAINS 4 TNFR-CYS REPEATS.

This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; AF104419; AAD03056.1; -
 DR EMBL; AF134240; AAD29688.1; -
 DR EMBL; AF217796; AAF35244.1; -
 DR EMBL; AF217793; AAF33685.1; -
 DR EMBL; AF217794; AAF33686.1; -
 DR EMBL; AL121845; CAC03688.1; -
 DR EMBL; BC017065; AAH17065.1; -
 DR GenBank; HGNC:11921; TNFRSF6B.
 DR MIN; 603361; -
 DR HSP; 014763; 1D0G.
 DR InterPro; IPR001368; TNFR_c6.
 DR Pfam; PF00020; TNFR_c6; 4.
 DR ProDom; PD000771; TNFR_c6; 1.
 DR SMART; SM00208; TNFR; 3.
 DR PROSITE; PS00652; TNFR_NGFR_1; 2.
 DR PROSITE; PS0050; TNFR_NGFR_2; 2.
 KW Receptor; Apoptosis; Glycoprotein; Repeat; Signal.
 FT SIGNAL 1 29
 FT CHAIN 30 300
 FT TUMOR NECROSIS FACTOR RECEPTOR
 FT SUPERFAMILY MEMBER 6B.
 FT REPEAT 31 70
 FT TNFR-CYS 1.
 FT REPEAT 72 113
 FT TNFR-CYS 2.
 FT REPEAT 115 150
 FT TNFR-CYS 3.
 FT REPEAT 152 193
 FT TNFR-CYS 4.
 FT DISULFID 49 62
 FT BY SIMILARITY.
 FT DISULFID 52 70
 FT BY SIMILARITY.
 FT DISULFID 73 88
 FT BY SIMILARITY.
 FT DISULFID 91 105
 FT BY SIMILARITY.
 FT DISULFID 95 113
 FT BY SIMILARITY.
 FT DISULFID 115 126
 FT BY SIMILARITY.
 FT DISULFID 132 150
 FT BY SIMILARITY.
 FT DISULFID 153 168
 FT BY SIMILARITY.
 FT DISULFID 174 193
 FT BY SIMILARITY.
 FT CARBOHYD 173 173
 FT N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 300 AA; 32679 MW; F90AEE33718449AF CRC64;

Query Match 0.8%; Score 8; DB 1; Length 300;
 Best Local Similarity 100.0%; Pred. No. 5.6;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 410 PCPPGTF 417
 DB 152 PCPPGTF 159

RESULT 6
 DCP2_PEA
 ID DCP2_PEA STANDARD; PRT; 405 AA.
 AC P51851;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update).
 DE Pyruvate decarboxylase isozyme 2 (EC 4.1.1.1) (PDC) (Fragment).
 GN PDC2.
 OS Pisum sativum (Garden pea).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
 OX NCBI_TAXID=3888;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. MIKO;
 RX MEDLINE=96215432; PubMed=8647075;
 RA Muecke U., Wohlfarth T., Fiedler U., Baeumlein H.,
 RA Ruecknagel K.P., Koenig S.;
 RT "Pyruvate decarboxylase from Pisum sativum. Properties, nucleotide

RT and amino acid sequences".
 RL Eur. J. Biochem. 237:373-382(1996).
 CC -!- CATALYTIC ACTIVITY: A 2-oxo acid - an aldehyde + CO(2).
 CC -!- COFACTOR: Binds 1 thiamine pyrophosphate and 1 metal ion per
 CC subunit.
 CC -!- SUBUNIT: HOMOOCTAMER (POTENTIAL).
 CC -!- SIMILARITY: BELONGS TO THE TPP ENZYMES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; Z66544; CAA91445.1; -
 DR HSP; P06672; 12PD.
 DR InterPro; IPR000399; TPP_enzyme.
 DR Pfam; PF00205; TPP_enzymes; 1.
 DR Pfam; PF02775; TPP_enzymes_C; 1.
 DR PROSITE; PS00187; TPP_ENZYMES; PARTIAL.
 KW Lyase; Decarboxylase; Flavoprotein; Thiamine pyrophosphate;
 KW Multigene family.
 FT NON_TER 1
 SQ SEQUENCE 405 AA; 44078 MW; 44DDCE90B38677FB CRC64;

Query Match 0.8%; Score 8; DB 1; Length 405;
 Best Local Similarity 100.0%; Pred. No. 7.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 KAVKPVLV 276
 DB 50 KAVKPVLV 57

RESULT 7

DCPL_TOBAC
 ID DCPL_TOBAC STANDARD; PRT; 418 AA.
 AC P51845;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Pyruvate decarboxylase isozyme 1 (EC 4.1.1.1) (PDC) (Fragment).
 GN PDC1.
 OS Nicotiana tabacum (Common tobacco).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
 OX NCBI_TAXID=4097;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Samsun; TISSUE=Leaf;
 RX MEDLINE=95375236; PubMed=7647304;
 RA Bucher M., Brander K., Sbicego S., Mandel T., Kuhlmeier C.;
 RT "Aerobic fermentation in tobacco pollen".
 RL Plant Mol. Biol. 28:739-750(1995).
 CC -!- CATALYTIC ACTIVITY: A 2-oxo acid - an aldehyde + CO(2).
 CC -!- COFACTOR: Binds 1 thiamine pyrophosphate and 1 metal ion per
 CC subunit.
 CC -!- SUBUNIT: HOMOOCTAMER (POTENTIAL).
 CC -!- TISSUE SPECIFICITY: LEAVES.
 CC -!- INDUCTION: ANAEROBICALLY.
 CC -!- SIMILARITY: BELONGS TO THE TPP ENZYMES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; X81854; CAA57447.1; -
DR HSP; P06169; IPVD.
DR InterPro; IPR000399; TPP_enzyme.
DR Pfam; PF0205; TPP_enzymes; 1.
DR Pfam; PF02775; TPP_enzymes; C; 1.
DR Pfam; PF02776; TPP_enzymes; N; 1.
DR PROSITE; PS00187; TPP_ENZYMES; PARTIAL.
DR Lysase; Decarboxylase; Flavoprotein; Thiamine pyrophosphate;
KW Multigene family.
FT NON_TER 1
FT NON_TER 418
SQ SEQUENCE 418 AA; 45010 MW; 7C3C33DAA9FFD804 CRC64;
Query Match 0.8%; Score 8; DB 1; Length 418;
Best Local Similarity 100.0%; Pred. No. 7.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 269 KAVKPVLV 276
Db 155 KAVKPVLV 162
RESULT 8
ID TRLT_HUMAN STANDARD; PRT; 430 AA.
AC Q96924; Q96JUI; Q9BUX7;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tumor necrosis factor receptor superfamily member TNFRSF19L precursor
DE (Receptor expressed in lymphoid tissues).
GN TNFRSF19L OR RELT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF N-TERMINUS.
RC TISSUE=Lymphoma;
RX MEDLINE=21213541; PubMed=11313261;
RA Sica G.L., Zhu G., Tamada K., Liu D., Ni J., Chen L.;
RT "RELt, a new member of the tumor necrosis factor receptor superfamily,
RT is selectively expressed in hematopoietic tissues and activates
RT transcription factor NF-kappaB.";
RL Blood 97:2702-2707(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Retinoblastoma;
RA Isorai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Nimomiya K., Iwayanagi T.;
RT "NEO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon, and Eye;
RA Strausberg R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 121-430 FROM N.A.
RC TISSUE=Spleen;
RA Jikuya H., Takano J., Nomura N., Kikuno R., Nagase T., Ohara O.;
RT "The nucleotide sequence of a long cDNA clone isolated from human
RT spleen.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Mediates activation of NF-kappa-B. May play a role in T-
CC cell activation.
CC -!- SUBUNIT: Associates with TRAF1.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Probable).

CC -!- TISSUE SPECIFICITY: Highest levels are in spleen, lymph node,
CC thymus, peripheral blood leukocytes, bone marrow and fetal liver.
CC Very low levels in skeletal muscle, testis and colon. Not detected
CC in brain, kidney and pancreas.
CC -!- SIMILARITY: CONTAINS 1 TNFR-CYS REPEAT.
CC -!- CAUTION: Ref.4 sequence differs from that shown due to several
CC frameshifts.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; AF319553; AAK77356.1; -
CC EMBL; AK027899; BAB55441.1; -
CC EMBL; BC001812; AAH01812.1; -
CC EMBL; BC017279; AAH17279.1; -
CC EMBL; AK074128; BAB84954.1; ALT_FRAME.
CC Genew; HGNC:13764; TNFRSF19L.
CC PROSITE; PS00652; TNFR_NGFR_1; FALSE_NEG.
CC PROSITE; PS00500; TNFR_NGFR_2; FALSE_NEG.
KW Receptor; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 26
FT CHAIN 27 430
FT DOMAIN 27 162 TUMOR NECROSIS FACTOR RECEPTOR
FT TRANSMEM 163 183 SUPERFAMILY MEMBER TNFRSF19L.
FT DOMAIN 184 430 EXTRACELLULAR (POTENTIAL).
FT REPEAT 50 90 CYTOPLASMIC (POTENTIAL).
FT DISULFID 51 65 TNFR-CYS.
FT DISULFID 71 90 BY SIMILARITY.
FT CARBOHYD 149 149 BY SIMILARITY.
FT CONFLICT 122 122 N-LINKED (GLCNAC...) (POTENTIAL).
FT CONFLICT 187 187 D -> S (IN REF. 4).
FT CONFLICT 273 273 K -> E (IN REF. 2).
FT CONFLICT 273 273 H -> R (IN REF. 2).
FT CONFLICT 379 380 DL -> TR (IN REF. 3; AAH01812).
SQ SEQUENCE 430 AA; 46092 MW; 4A5AB9AE32D36101 CRC64;
Query Match 0.8%; Score 8; DB 1; Length 430;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 410 PCPPGTFS 417
Db 50 PCPPGTFS 57
RESULT 9
ID TRLT_MACFA STANDARD; PRT; 430 AA.
AC Q9N092;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tumor necrosis factor receptor superfamily member TNFRSF19L precursor
DE (Receptor expressed in lymphoid tissues).
GN TNFRSF19L OR RELT.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=21458551; PubMed=11574149;
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirata M., Suto Y.,
RA Hirai M., Terao K., Suzuki Y., Sugano S., Hashimoto K., Kusuda J.;
RT "Assignment of 118 novel cDNAs of cynomolgus monkey brain to human
RT chromosomes.";

RL Gene 275:31-37(2001).
CC -!- FUNCTION: Mediates activation of NF-kappa-B (By similarity). May
CC play a role in T-cell activation.
CC -!- SUBUNIT: Associates with TRAF1 (By similarity).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC -!- SIMILARITY: CONTAINS 1 TNFR-CYS REPEAT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AB046039; BAB01621.1; -
DR InterPro: IPR001368; TNFR_C6
DR PROSITE: PS00652; TNFR_NGFR_1; FALSE_NEG.
DR PROSITE: PS00050; TNFR_NGFR_2; FALSE_NEG.
DR SMART: SMO0208; TNFR; 1.
KW Receptor; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 26 BY SIMILARITY.
FT CHAIN 27 430 TUMOR NECROSIS FACTOR RECEPTOR
FT SUPERFAMILY MEMBER TNFRSF19L.
FT DOMAIN 27 162 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 163 183 POTENTIAL.
FT DOMAIN 184 430 CYTOPLASMIC (POTENTIAL).
FT REPEAT 50 90 TNFR-CYS.
FT DISULFID 51 65 BY SIMILARITY.
FT DISULFID 71 90 BY SIMILARITY.
FT CARBOHYD 149 149 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 430 AA: 45850 MW; BABDE92593E1E859 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 430;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 410 PCPPGTFS 417
DB 50 PCPPGTFS 57
IIIIIIII

RESULT 10
DCP3_ORYSA STANDARD; PRT; 585 AA.
AC P51849;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pyruvate decarboxylase isozyme 3 (EC 4.1.1.1) (PDC).
GN PDC3.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Indica-IR54; TISSUE=Callus;
RA Hossain M.A., McGee J.D., Grover A., Dennis E., Peacock W.J.,
RA Hodges T.K.;
RT "Nucleotide sequence of a rice genomic pyruvate decarboxylase gene
RT that lacks introns: a pseudo-gene";
RL Plant Physiol. 106:1697-1698(1994).
CC -!- CATALYTIC ACTIVITY: A 2-oxo acid = an aldehyde + CO(2).
CC -!- COFACTOR: Binds 1 thiamine pyrophosphate and 1 metal ion per
CC subunit.
CC -!- SUBUNIT: HOMOOCTAMER (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE TPP ENZYMES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U07338; AAA68289.1; -
DR HSP; P06169; LPVD.
DR InterPro: IPR000399; TPP_enzyme.
DR Pfam: PF00205; TPP_enzymes_C; 1.
DR Pfam: PF02775; TPP_enzymes_N; 1.
DR PROSITE: PS00187; TPP_ENZYMES; FALSE_NEG.
KW Lyase; Decarboxylase; Flavoprotein; Thiamine pyrophosphate;
KW Multigene family.
FT ACT_SITE 71 71 BY SIMILARITY.
SQ SEQUENCE 585 AA: 62456 MW; F26AFA7A5EC8AAE52 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 585;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 269 KAVKPVLV 276
DB 231 KAVKPVLV 238
IIIIIIII

RESULT 11
DCP2_ORYSA STANDARD; PRT; 603 AA.
ID DCP2_ORYSA STANDARD; PRT; 603 AA.
AC P51848;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pyruvate decarboxylase isozyme 2 (EC 4.1.1.1) (PDC).
GN PDC2.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Indica-IR54; TISSUE=Callus;
RA Huq M.E., Hossain M.A., Hodges T.K.;
RA "Cloning and sequencing of a cDNA encoding pyruvate decarboxylase 2
RA gene from rice";
RL (In) Plant Gene Register PGR95-072.
CC -!- CATALYTIC ACTIVITY: A 2-oxo acid = an aldehyde + CO(2).
CC -!- COFACTOR: Binds 1 thiamine pyrophosphate and 1 metal ion per
CC subunit.
CC -!- SUBUNIT: HOMOOCTAMER (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE TPP ENZYMES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U37350; AAA90948.1; -
DR EMBL: U38199; AAB40530.1; -
DR HSP; P06169; LPVD.
DR InterPro: IPR000399; TPP_enzyme.
DR Pfam: PF00205; TPP_enzymes_C; 1.
DR Pfam: PF02775; TPP_enzymes_N; 1.
DR PROSITE: PS00187; TPP_ENZYMES; FALSE_NEG.
KW Lyase; Decarboxylase; Flavoprotein; Thiamine pyrophosphate;
KW Multigene family.
FT ACT_SITE 89 89 BY SIMILARITY.

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U07338; AAA68289.1; -
DR HSP; P06169; LPVD.
DR InterPro: IPR000399; TPP_enzyme.
DR Pfam: PF00205; TPP_enzymes_C; 1.
DR Pfam: PF02775; TPP_enzymes_N; 1.
DR PROSITE: PS00187; TPP_ENZYMES; FALSE_NEG.
KW Lyase; Decarboxylase; Flavoprotein; Thiamine pyrophosphate;
KW Multigene family.
FT ACT_SITE 71 71 BY SIMILARITY.
SQ SEQUENCE 585 AA: 62456 MW; F26AFA7A5EC8AAE52 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 585;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 269 KAVKPVLV 276
DB 231 KAVKPVLV 238
IIIIIIII

RESULT 11
DCP2_ORYSA STANDARD; PRT; 603 AA.
ID DCP2_ORYSA STANDARD; PRT; 603 AA.
AC P51848;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pyruvate decarboxylase isozyme 2 (EC 4.1.1.1) (PDC).
GN PDC2.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Indica-IR54; TISSUE=Callus;
RA Huq M.E., Hossain M.A., Hodges T.K.;
RA "Cloning and sequencing of a cDNA encoding pyruvate decarboxylase 2
RA gene from rice";
RL (In) Plant Gene Register PGR95-072.
CC -!- CATALYTIC ACTIVITY: A 2-oxo acid = an aldehyde + CO(2).
CC -!- COFACTOR: Binds 1 thiamine pyrophosphate and 1 metal ion per
CC subunit.
CC -!- SUBUNIT: HOMOOCTAMER (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE TPP ENZYMES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U37350; AAA90948.1; -
DR EMBL: U38199; AAB40530.1; -
DR HSP; P06169; LPVD.
DR InterPro: IPR000399; TPP_enzyme.
DR Pfam: PF00205; TPP_enzymes_C; 1.
DR Pfam: PF02775; TPP_enzymes_N; 1.
DR PROSITE: PS00187; TPP_ENZYMES; FALSE_NEG.
KW Lyase; Decarboxylase; Flavoprotein; Thiamine pyrophosphate;
KW Multigene family.
FT ACT_SITE 89 89 BY SIMILARITY.

SQ SEQUENCE 603 AA; 64301 MW; 65B83A6468DCC2EB CRC64;

Query Match 0.8%; Score 8; DB 1; Length 603;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 269 KAVKPKVLV 276
DB 249 KAVKPKVLV 256

RESULT 12
KRP2_RAT

ID KRP2_RAT STANDARD; PRT; 671 AA.

AC Q62909;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Kinesin-related protein 2.
GN KRP2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Testis;
RX MEDLINE=96228687; PubMed=8688559;
RA Sperry A.O.; Zhao L.-P.;
RT "Kinesin-related proteins in the mammalian testes: candidate motors
for meiosis and morphogenesis";
RL Mol. Biol. Cell 7:289-305(1996).
RN [2]
RP REVISIONS.
RA Sperry A.O.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MOTOR PROTEIN, ACTIVE IN MEIOSIS.
CC -!- TISSUE SPECIFICITY: TESTIS. LOCALIZED TO THE MEIOGONIACTIVELY ACTIVE
CELLS OF THE SEMINIFEROUS EPITHELIA IN THE TESTIS.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. MCAK/KIF2
SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

CC EMBL; U44979; AAC33528.1; -
CC HSP; P17119; 3KAR.
CC InterPro; IPR001752; kinesin_motor.
CC Pfam; PF00225; kinesin; 1.
CC PRINTS; PR00380; KINESINHEAVY.
CC SMART; SM00129; KISC; 1.
CC PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
CC PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
CC Motor protein; Microtubules; ATP-binding; Coiled coil;
KW Nuclear protein.
KW DOMAIN 1 198 GLOBULAR (POTENTIAL).
FT DOMAIN 199 540 KINESIN-MOTOR (BY SIMILARITY).
FT DOMAIN 541 671 COILED COIL (POTENTIAL).
FT NP_BIND 294 301 ATP (POTENTIAL).
SQ SEQUENCE 671 AA; 75661 MW; F2B54598C78DE8DE CRC64;

Query Match 0.8%; Score 8; DB 1; Length 671;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 846 VISVPSKC 853
DB 226 VISVPSKC 233

RESULT 13
MCAK_CRIGR

ID MCAK_CRIGR STANDARD; PRT; 718 AA.

AC P70096;
DT 01-NOV-1997 (Rel. 35, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Mitotic centromere-associated kinesin (MCAK) (Kinesin-like protein 6).
GN KNSL6.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetus.
OX NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95122643; PubMed=7822426;
RA Wordeman L.; Mitchison T.J.;
RT "Identification and partial characterization of mitotic centromere-
associated kinesin, a kinesin-related protein that associates with
centromeres during mitosis";
RL J. Cell Biol. 128:95-105(1995).
RN [2]
RP REVISIONS.
RA Wordeman L.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PRESENT THROUGHOUT THE CELL CYCLE, ASSOCIATES WITH THE
CENTROMERES AT EARLY PROPHASE, AND REMAINS ASSOCIATED WITH THE
CENTROMERE UNTIL AFTER TELOPHASE.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. MCAK/KIF2
SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

CC EMBL; U11790; AAB17358.2; -
CC HSP; P17119; 3KAR.
CC InterPro; IPR001752; kinesin_motor.
CC Pfam; PF00225; kinesin; 1.
CC PRINTS; PR00380; KINESINHEAVY.
CC SMART; SM00129; KISC; 1.
CC PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
CC PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
CC Motor protein; Microtubules; ATP-binding; Coiled coil;
KW Nuclear protein.
KW DOMAIN 1 246 GLOBULAR (POTENTIAL).
FT DOMAIN 247 612 KINESIN-MOTOR (BY SIMILARITY).
FT DOMAIN 613 651 COILED COIL (POTENTIAL).
FT DOMAIN 689 716 COILED COIL (POTENTIAL).
FT NP_BIND 342 349 ATP (POTENTIAL).
FT DOMAIN 409 412 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
SQ SEQUENCE 718 AA; 80918 MW; 16ABD8BC66AD11B2 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 718;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 846 VISVPSKC 853
DB 274 VISVPSKC 281

RESULT 14
TAT_SIVCZ

ID TAT_SIVCZ STANDARD; PRT; 100 AA.
AC P17285;

```
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE TAT protein (Transactivating regulatory protein).
GN TAT.
OS Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11723;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90259077; PubMed=2188136;
RA Huet T., Cheynier R., Meyerhans A., Roelants G., Wain-Hobson S.;
RT "Genetic organization of a chimpanzee lentivirus related to HIV-1.";
RL Nature 345:356-359(1990).
CC -!- FUNCTION: TRANSCRIPTIONAL REGULATOR THAT ACTS BY BINDING TO THE
CC TRANS-ACTIVATING RESPONSIVE SEQUENCE (TAR) RNA ELEMENT AND
CC ACTIVATES TRANSCRIPTION INITIATION AND/OR ELONGATION FROM THE LTR
CC PROMOTER.
CC -!- SUBUNIT: BINDS CYCLIN T1 (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52154; CA336404.1; -
DR PIR; S09987; TNLJSI.
DR HIV; X52154; TATSCPZ.
DR InterPro: IPR001831; HIV_Tat.
DR Pfam; PF00539; Tat; 1.
DR PRINTS; PR00055; HIVTATDOMAIN.
KW Transcription regulation; Activator; RNA-binding; Nuclear protein;
KW AIDS.
SQ SEQUENCE 100 AA; 11209 MW; 1B78830B90EED50E CRC64;

Query Match 0.7%; Score 7; DB 1; Length 100;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 658 GSKNOD 664 .
DB 62 GSKNOD 68

RESULT 15
RK24_GUITH
ID RK24_GUITH STANDARD; PRT; 101 AA.
AC O46905;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Chloroplast 50S ribosomal protein L24.
GN RPL24.
OS Guillardia theta (Cryptomonas phi).
OC Chloroplast.
OC Eukaryota; Cryptophyta; Cryptomonadaceae; Guillardia.
OX NCBI_TaxID=55529;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97283757; PubMed=9137835;
RA Wang S.L., Liu X.-Q., Douglas S.E.;
RT "The large ribosomal protein gene cluster of a cryptomonad plastid:
RT gene organization, sequence and evolutionary implications.";
RL Biochem. Mol. Biol. Int. 41:1035-1044(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99128221; PubMed=9929392;
RA Douglas S.E., Penny S.L.;
RT "The plastid genome of the cryptophyte alga, Guillardia theta;
```

```
RT complete sequence and conserved syntenic groups confirm its common
RT ancestry with red algae.";
RL J. Mol. Evol. 48:236-244(1999).
CC -!- SIMILARITY: BELONGS TO THE L24P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF041468; AAC35714.1; -
DR InterPro: IPR000302; KOW_motif.
DR InterPro: IPR003256; Ribosomal_L24.
DR Pfam; PF00467; KOW; 1.
DR ProDom; PD001677; Ribosomal_L24; 1.
DR TIGRFAMs; TIGR01079; rplX_bact; 1.
DR PROSITE; PS01108; RIBOSOMAL_L24; FALSE_NEG.
KW Ribosomal protein; Chloroplast.
SQ SEQUENCE 101 AA; 11572 MW; 306E2E229C636131 CRC64;

Query Match 0.7%; Score 7; DB 1; Length 101;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 792 LKNINIK 798
DB 35 LKNINIK 41
```

Search completed: May 12, 2003, 13:39:34
Job time : 17.0658 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 07:48:38 ; Search time 46.4563 Seconds
(without alignments)
2762.174 Million cell updates/sec

Title: US-10-073-333A-2
Perfect score: 5357
Sequence: 1 MLFRARGVGRGWRPAEA.....TCYFWKKKKKKTLNLFN 963

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

- 1: /SID22/gcgdata/geneseq/geneseq-embl/AA1980.DAT.*
- 2: /SID22/gcgdata/geneseq/geneseq-embl/AA1981.DAT.*
- 3: /SID22/gcgdata/geneseq/geneseq-embl/AA1982.DAT.*
- 4: /SID22/gcgdata/geneseq/geneseq-embl/AA1983.DAT.*
- 5: /SID22/gcgdata/geneseq/geneseq-embl/AA1984.DAT.*
- 6: /SID22/gcgdata/geneseq/geneseq-embl/AA1985.DAT.*
- 7: /SID22/gcgdata/geneseq/geneseq-embl/AA1986.DAT.*
- 8: /SID22/gcgdata/geneseq/geneseq-embl/AA1987.DAT.*
- 9: /SID22/gcgdata/geneseq/geneseq-embl/AA1988.DAT.*
- 10: /SID22/gcgdata/geneseq/geneseq-embl/AA1989.DAT.*
- 11: /SID22/gcgdata/geneseq/geneseq-embl/AA1990.DAT.*
- 12: /SID22/gcgdata/geneseq/geneseq-embl/AA1991.DAT.*
- 13: /SID22/gcgdata/geneseq/geneseq-embl/AA1992.DAT.*
- 14: /SID22/gcgdata/geneseq/geneseq-embl/AA1993.DAT.*
- 15: /SID22/gcgdata/geneseq/geneseq-embl/AA1994.DAT.*
- 16: /SID22/gcgdata/geneseq/geneseq-embl/AA1995.DAT.*
- 17: /SID22/gcgdata/geneseq/geneseq-embl/AA1996.DAT.*
- 18: /SID22/gcgdata/geneseq/geneseq-embl/AA1997.DAT.*
- 19: /SID22/gcgdata/geneseq/geneseq-embl/AA1998.DAT.*
- 20: /SID22/gcgdata/geneseq/geneseq-embl/AA1999.DAT.*
- 21: /SID22/gcgdata/geneseq/geneseq-embl/AA2000.DAT.*
- 22: /SID22/gcgdata/geneseq/geneseq-embl/AA2001.DAT.*
- 23: /SID22/gcgdata/geneseq/geneseq-embl/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5357	100.0	963	22 AAB70255	TR16-short receptor
2	5307	99.1	1027	22 AAB70256	TR16-long receptor
3	2821.5	52.7	1013	21 AAB26179	Human CASB619 prot
4	2815.5	52.6	1001	22 AAB35333	Human TR13 recepto
5	2811.5	52.5	1013	22 AAB33845	Amino acid sequenc
6	2796	52.2	1013	22 AAU12190	Human PRO4985 poly
7	2712.5	50.6	911	22 AAB33850	Amino acid sequenc
8	2702.5	50.4	870	22 AAB33851	Amino acid sequenc
9	2526	47.2	464	22 AAB48377	Human SEC10 protei
10	2222	41.5	411	22 AAB48372	Human SEC5 protein

11	2000	37.3	750	22 AAB35328	Human TR13 recepto
12	1258.5	23.5	495	20 AAY59972	Human endometrium
13	1208.5	22.6	383	22 AAB83853	Amino acid sequenc
14	883	16.5	372	22 AAB85768	Human seven-transm
15	705.5	13.2	209	22 AAB83852	Amino acid sequenc
16	452	8.4	81	22 ABB39918	Peptide #7424 enco
17	452	8.4	81	22 ABB24471	Protein #6470 enco
18	452	8.4	81	22 AAM60663	Human brain expres
19	452	8.4	81	22 AAM73335	Human bone marrow
20	452	8.4	81	22 AAM33535	Human bone marrow
21	452	8.4	81	22 AAM33535	Peptide #7572 enco
22	444.5	8.3	208	21 AAB53442	Human peptidase enco
23	442	8.3	74	22 AAB70281	Human colon cancer
24	370.5	6.9	147	22 AAB83849	Peptide #25. Unid
25	353	6.6	60	22 ABB38686	Peptide fragment o
26	353	6.6	60	22 ABB38686	Peptide #6192 enco
27	353	6.6	60	22 ABB23759	Protein #5758 enco
28	353	6.6	60	22 AAM59318	Human brain expres
29	353	6.6	60	22 AAM32149	Human bone marrow
30	353	6.6	60	22 ABB41680	Peptide #6186 enco
31	353	6.6	64	22 AAM72925	Human peptidase enco
32	353	6.6	64	22 ABB42760	Human bone marrow
33	350	6.5	150	20 AAY12274	Human peptidase enco
34	263	4.9	105	21 AAB26180	Human CASB619 prot
35	205	3.8	1605	21 AAB19805	Mouse laminin 2 ga
36	205	3.8	1605	21 AAB48454	Mouse laminin 2 ga
37	205	3.8	1605	21 ABB81596	Mouse laminin 10 t
38	200	3.7	3635	23 ABB81589	Mouse laminin 10 t
39	200	3.7	3635	23 AAM50357	Mouse laminin-15 a
40	197.5	3.7	1572	21 AAB19806	Mouse laminin 2 ma
41	197.5	3.7	1572	21 AAB48455	Mouse laminin 8 po
42	197.5	3.7	1572	21 ABB81597	Mouse laminin 10 t
43	188.5	3.5	999	23 ABB80926	Human breast cance
44	188.5	3.5	999	23 ABB61893	Prostate cancer-as
45	182	3.4	2931	22 ABB68229	Drosophila melanog

ALIGNMENTS

RESULT 1

AAB70255

ID AAB70255 standard; protein; 963 AA.

XX

AC AAB70255;

XX

DT 10-MAY-2001 (first entry)

XX

DE TR16-short receptor protein.

XX

DE TR16 receptor; tumour necrosis factor receptor superfamily;

KW apoptosis; inflammatory; cancer; immune; neurodegenerative.

KW

XX

OS Unidentified.

XX

PN WO200112671-A1.

XX

PD 22-FEB-2001.

XX

PF 10-AUG-2000; 2000WO-US21885.

XX

PR 12-AUG-1999; 99US-0148348.

PR

PR 13-AUG-1999; 99US-0148683.

PR

PR 13-AUG-1999; 99US-0148870.

PR

PR 16-AUG-1999; 99US-0148758.

PR

PR 17-AUG-1999; 99US-0149181.

PR

PR 18-AUG-1999; 99US-0149453.

PR

PR 19-AUG-1999; 99US-0149498.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

XX Ruben SM, Young PE, Baker KP.

PI

XX

DR WPI: 2001-138754/14.
XX New nucleic acid molecule encoding a TR16 tumor necrosis factor
PT receptor polypeptide, useful for the diagnosis and treatment of cancer,
PT autoimmune disorders and cardiovascular diseases -
XX Claim 1; Fig 1; 286pp; English.
PS The present invention relates to a TR16 receptor (tumour necrosis
CC factor receptor superfamily). The invention is useful treating
CC diseases and disorders associated with the inhibited or increased
CC apoptosis. In particular inflammatory diseases, cancers, immune and
CC neurodegenerative disorders may be treated.
XX Sequence 963 AA:
SQ Query Match 100.0%; Score 5357; DB 22; Length 963;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 963; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MLFRAGPVVRGWRGPAEAPRRGRSPWPSPAWICWALAGCAAWAGDLPSSSSRPLPP 60
Db 1 MLFRAGPVVRGWRGPAEAPRRGRSPWPSPAWICWALAGCAAWAGDLPSSSSRPLPP 60
Qy 61 COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFCASGEYLEMKNQVC 120
Db 61 COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFCASGEYLEMKNQVC 120
Qy 121 SKCGEYSLGSGIKFDEWDELPAGFSNATFMDTVGSDSRPDGNCNNSWIPRGNYIE 180
Db 121 SKCGEYSLGSGIKFDEWDELPAGFSNATFMDTVGSDSRPDGNCNNSWIPRGNYIE 180
Qy 181 SNRDDCTVSLIYAVHLKSGYVFEFFQYVDNNIFFEFFIQNDQCEMDTTTDDKWKLTDN 240
Db 181 SNRDDCTVSLIYAVHLKSGYVFEFFQYVDNNIFFEFFIQNDQCEMDTTTDDKWKLTDN 240
Qy 241 GWSHVMKSGTNILYWRRTGILMGSKAVPVLVKNITIEGVAYTSECFCKPGTFSN 300
Db 241 GWSHVMKSGTNILYWRRTGILMGSKAVPVLVKNITIEGVAYTSECFCKPGTFSN 300
Qy 301 KPGSNQVCPRTYSEKAKECIRCKDQSOFSGSECTERPCTTKDYFOIHTPCDEEG 360
Db 301 KPGSNQVCPRTYSEKAKECIRCKDQSOFSGSECTERPCTTKDYFOIHTPCDEEG 360
Qy 361 KTOIMYKWIETPKICREDLTDAILRPPSGEKDKPCPNPGFYNNSSCHPCPPGTFSDGT 420
Db 361 KTOIMYKWIETPKICREDLTDAILRPPSGEKDKPCPNPGFYNNSSCHPCPPGTFSDGT 420
Qy 421 KECRCPAGTEPALGFYKWNVLFGNMKTSFNVGNSKCDGMNGWEVAGDHIQSGAGS 480
Db 421 KECRCPAGTEPALGFYKWNVLFGNMKTSFNVGNSKCDGMNGWEVAGDHIQSGAGS 480
Qy 481 DNDYLILNLHIFGFPPTSMTCATGSELGRITFVFETLCSADCVLYFMVDINRKSTNVVE 540
Db 481 DNDYLILNLHIFGFPPTSMTCATGSELGRITFVFETLCSADCVLYFMVDINRKSTNVVE 540
Qy 541 SNGGTKEQAYTHIIFKNATFTTFAFORTNQDQNRREINDVVKIYSITATNVDGVAS 600
Db 541 SNGGTKEQAYTHIIFKNATFTTFAFORTNQDQNRREINDVVKIYSITATNVDGVAS 600
Qy 601 SCRACALGSESGSCVCPGPGHYTEKETNQCKECPDPTYSIHQVYGEACIPGPGSK 660
Db 601 SCRACALGSESGSCVCPGPGHYTEKETNQCKECPDPTYSIHQVYGEACIPGPGSK 660
Qy 661 NNQDSVCYSDGCFYHEKENQILHVDNFSLSSVGLMNGPSTSKGTKYHFENISLCGH 720
Db 661 NNQDSVCYSDGCFYHEKENQILHVDNFSLSSVGLMNGPSTSKGTKYHFENISLCGH 720
Qy 721 EGKMKALCTNNITDFTVKEIVAGSDDDYTNLGAFCVQSTIIPSESKGFRALSSQSIILA 780
Db 721 EGKMKALCTNNITDFTVKEIVAGSDDDYTNLGAFCVQSTIIPSESKGFRALSSQSIILA 780
Qy 781 DTFIGVTETTLKINIKEDMPVPTSQIPDVHFFYKSTATTSCINGRSTAVKMCNPT 840

Db 781 DTFIGVTETTLKINIKEDMPVPTSQIPDVHFFYKSTATTSCINGRSTAVKMCNPT 840
Qy 841 KSGAGVISVPSKCPAGTCDGCTFFFLWESAEACPLCTEHDFHEIEGACKRGFQETLYVWN 900
Db 841 KSGAGVISVPSKCPAGTCDGCTFFFLWESAEACPLCTEHDFHEIEGACKRGFQETLYVWN 900
Qy 901 EPKWCIKIGISLPEKKLATCETVDFWLKVGAGVGAFATVLLVALTCYFWKKKOKKKTTILN 960
Db 901 EPKWCIKIGISLPEKKLATCETVDFWLKVGAGVGAFATVLLVALTCYFWKKKOKKKTTILN 960
Qy 961 LPN 963
Db 961 LPN 963
RESULT 2
AAB70256
ID AAB70256 standard; protein: 1027 AA.
XX AC AAB70256;
XX DT 10-MAY-2001 (first entry)
XX DE TR16-long receptor protein.
XX KW TR16-receptor; tumour necrosis factor receptor superfamily;
XX OS Unidentified.
XX PN WO200112671-A1.
XX PD 22-FEB-2001.
XX PF 10-AUG-2000; 2000WO-US21885.
XX PR 12-AUG-1999; 99US-0148348.
XX PR 13-AUG-1999; 99US-0148683.
XX PR 13-AUG-1999; 99US-0148870.
XX PR 16-AUG-1999; 99US-0148758.
XX PR 17-AUG-1999; 99US-0149181.
XX PR 18-AUG-1999; 99US-0149453.
XX PR 19-AUG-1999; 99US-0149498.
XX (HUMA-) HUMAN GENOME SCI INC.
XX PI Ruben SM, Young PE, Baker KP;
XX WPI: 2001-138754/14.
XX New nucleic acid molecule encoding a TR16 tumor necrosis factor
PT receptor polypeptide, useful for the diagnosis and treatment of cancer,
PT autoimmune disorders and cardiovascular diseases -
XX Disclosure: Fig 4; 286pp; English.
XX The present invention relates to a TR16 receptor (tumour necrosis
CC factor receptor superfamily). The invention is useful treating
CC diseases and disorders associated with the inhibited or increased
CC apoptosis. In particular inflammatory diseases, cancers, immune and
CC neurodegenerative disorders may be treated.
XX Sequence 1027 AA;
SQ Query Match 99.1%; Score 5307; DB 22; Length 1027;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 953; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MLFRAGPVVRGWRGPAEAPRRGRSPWPSPAWICWALAGCAAWAGDLPSSSSRPLPP 60
Db 1 MLFRAGPVVRGWRGPAEAPRRGRSPWPSPAWICWALAGCAAWAGDLPSSSSRPLPP 60

```
QY 61 CQEKDYHFEYTECDSSGRWRVAIPNSAVDCSLPDPVRGKECTFCSCASGEYLEMKNQVC 120
Db 61 CQEKDYHFEYTECDSSGRWRVAIPNSAVDCSLPDPVRGKECTFCSCASGEYLEMKNQVC 120
QY 121 SKCGEGTSLGSGIKDEWDELPAFESNIATFMDTVVGPDSRDCNNSWIPRGNYIE 180
Db 121 SKCGEGTSLGSGIKDEWDELPAFESNIATFMDTVVGPDSRDCNNSWIPRGNYIE 180
QY 181 SNRDDCTVSLIYAVHLKSGYVFEFYQYVDNNIFFEFFIIONDQCQEMDTTDDKWKLTND 240
Db 181 SNRDDCTVSLIYAVHLKSGYVFEFYQYVDNNIFFEFFIIONDQCQEMDTTDDKWKLTND 240
QY 241 GENGSHVMLKSGTNILYRTTIGLMGSKAVKPVLVKNITIEGVAYTSECFPCPKPTFSN 300
Db 241 GENGSHVMLKSGTNILYRTTIGLMGSKAVKPVLVKNITIEGVAYTSECFPCPKPTFSN 300
QY 301 KPGSFNCOVCPRNTYSEKGAKEICRCKDDSOFSGSSSECTERPPCTTKDYFIHTPCDEEG 360
Db 301 KPGSFNCOVCPRNTYSEKGAKEICRCKDDSOFSGSSSECTERPPCTTKDYFIHTPCDEEG 360
QY 361 KTOIMYKWIPEKICREDLTDAILRPPSGEKKDCPPCNPGFYNNGSSSCHPCPPGTSDGT 420
Db 361 KTOIMYKWIPEKICREDLTDAILRPPSGEKKDCPPCNPGFYNNGSSSCHPCPPGTSDGT 420
QY 421 KECRCPAGTEPALGFEYKWNVLPNGMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGS 480
Db 421 KECRCPAGTEPALGFEYKWNVLPNGMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGS 480
QY 481 DNDYLILNLHPIGFKPPTMTGATGSELGRIETVFETLCSADCVLFWVDINKRSTNVVE 540
Db 481 DNDYLILNLHPIGFKPPTMTGATGSELGRIETVFETLCSADCVLFWVDINKRSTNVVE 540
QY 541 SWGGTKEKQAYTHIFKNATFTFTFAFORTNOQDNRFRINDMVKIYSITATNAVGVAS 600
Db 541 SWGGTKEKQAYTHIFKNATFTFTFAFORTNOQDNRFRINDMVKIYSITATNAVGVAS 600
QY 601 SCRACALGSEQSSGVCPGPHYIEKETNQCCEKCPDDYLSLTHQVYGEACIPCGPGSK 660
Db 601 SCRACALGSEQSSGVCPGPHYIEKETNQCCEKCPDDYLSLTHQVYGEACIPCGPGSK 660
QY 661 NNQDHSVCSYSDCFYHEKENQILHYDFSNLSSVGLMNGPSTSKGTYPHFENISLCGH 720
Db 661 NNQDHSVCSYSDCFYHEKENQILHYDFSNLSSVGLMNGPSTSKGTYPHFENISLCGH 720
QY 721 EGKMMALCTNNITDFTVKEIVAGSDDTYNLVGAFVCSQSTIIPSESGFRALSSQSIIILA 780
Db 721 EGKMMALCTNNITDFTVKEIVAGSDDTYNLVGAFVCSQSTIIPSESGFRALSSQSIIILA 780
QY 781 DTFIGYVETTLKNNIKEDMPVPPTSQIPDVHFFYKSSATTSCINGRSTAVKMRCPNT 840
Db 781 DTFIGYVETTLKNNIKEDMPVPPTSQIPDVHFFYKSSATTSCINGRSTAVKMRCPNT 840
QY 841 KSGAGVISVPSKPCAGTCDCGTFYFLWESAECPLCTEHDFHIEGACKRGFOETLYVWN 900
Db 841 KSGAGVISVPSKPCAGTCDCGTFYFLWESAECPLCTEHDFHIEGACKRGFOETLYVWN 900
QY 901 EPKWCIGISLPEKKLATCETVDLWLVKGVAGGAFVALLVALTCYFWKKNQK 953
Db 901 EPKWCIGISLPEKKLATCETVDLWLVKGVAGGAFVALLVALTCYFWKKNQK 953
```

RESULT 3

```
AAB26179
ID AAB26179 standard; Protein; 1013 AA.
XX
AC AAB26179;
XX
DT 12-FEB-2001 (first entry)
XX
DE Human CASB619 protein #1.
XX
KW Human; CASB619; cancer; autoimmune disease; immunogen; vaccine;
KW epitope.
```

```
XX Homo sapiens.
OS WO200058460-A2.
PN 05-OCT-2000.
PD 20-MAR-2000; 2000WO-EP02478.
PF 26-MAR-1999; 99GB-0007113.
PR 25-SEP-1999; 99GB-0022858.
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX Bruck CEM, Cassart J, Coche T, Vinals De Bassols YC;
PI WPI; 2000-864923/64.
DR N-PSDB; AAA95442.
XX Novel CASB619 polypeptides useful for diagnosing, and as vaccines for
PT prophylactic and therapeutic treatment of, cancers, particularly
PT ovarian and colon carcinoma, and autoimmune diseases
XX Claim 4; Page 54-56; 68pp: English.
XX The present sequence comprises the human CASB619 protein sequence. This
CC protein is thought to be specifically or over-expressed in tumour cells,
CC and so can be used as a target for antigen-specific immune responses
CC which can cause destruction of the tumour cell. In addition, the protein
CC and gene can be used in cancer diagnosis, in the treatment of autoimmune
CC diseases and in vaccines against cancer and autoimmune disease. The
CC invention provides a number of epitopes derived from the protein which
CC can be used as immunogens.
XX Sequence 1013 AA;
```

```
Query Match 52.7%; Score 2821.5; DB 21; Length 1013;
Best Local Similarity 53.0%; Pred. No. 3.8e-214;
Matches 502; Conservative 161; Mismatches 260; Indels 25; Gaps 12;
QY 23 RGRSPWSPANICWALAGCAAAG---DLPSSSRPLPPCOEKDYHFEYTECDSSGR 79
Db 15 RGRTERIPR---LWRL---LLWAGTAFQVQTGTGELHACKSEYHYEYACDSTGSR 67
QY 80 WRVAIPNSAVDCSLPDPVRGKECTFCSCASGEYLEMKNQVCSKCGEYTSILGSGIKRDEW 139
Db 68 WRVAVPHTPGCLCTSLPDPVRGKECTFCSCNAGEFLDMKQSKPCAGRYSLGTGIRDEW 127
QY 140 DELPAGFSNIATFMDTVVGPDSRDCNNSWIPRGNYIESNRDCTVSLIYAVHLKKS 199
Db 128 DELPHGFASLSANMELDDSAES-TGNCITSSKWPVPGDYIASNTDECTATLMYAVNLKQS 186
QY 200 GYVFFEYQYVDNNIFFEFFIIONDQCQEMDTTDDKWKLTNDNGENGSHVMLKSGTNILYV 259
Db 187 GTVNFYFYYPDSSIIIEFFVQNDQCP-NADDSRWKMTTEK-WEFHSVELNNGNNVLYW 244
QY 260 RTTGILMGSKAVKPVLVKNITIEGVAYTSECFPCPKPTFSNKGSCNVCQCPRTYSEK 319
Db 245 RTTAFSVTKVPKPVLVNRNITAITGVAYTSECFPCPKPTGYADKQGSFCKLCIPANSYSNK 304
QY 320 AKECIRCKDDSQFS--GSSECTERPPCTTKDYFIHTPCDEEGKTOIMYKWIPEKICRED 377
Db 305 ETSCHOC-DPDKYSBKSSSCNVRPACTDKDYFTHYTTACDANGETQLMYKWAAPKICSED 363
QY 378 LTDAILRPPSGEKKDCPPCNPGFYNNGSSSCHPCPPGTSDGTKECRPCPAGTEPALGFE 437
Db 364 LEGAVKLPASGVKTHCPNCPNPGFFKTNSTQCPCPYGSYNGS-DCTRCPCAGTEPAVGFE 422
QY 438 YKWNVNLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGSNDYLLNLHPIGFKPP 497
Db 423 YKWNVNLPTNMETVLUSGINFEYKGMTGEVAGDHIYTAAGASDNDFMILTLLVPGFRPP 482
QY 498 TS-MTGATGSELGRITFVFETLCSADCVLYFMVDINRKNSTNVVESWGGTKEKQAYTHIF 556
```

Db 483 QSVMDTENKEVARITVFETLCSVNCLELYFMVGNSTNTPVETWKGSKQSYTYIE 542
Qy 557 KNATFTFAFORTNQODNRRFINDMWKYSITATNAVDGAVSSCRACALGSEQSGSSC 616
Db 543 ENTTSFTWAFORTTFHEASRYTNDVAKIYSINVTNVMNGVASCRCALASDVGSSC 602
Qy 617 VPCPPGHVIEKETNOCKECPDPTYLSIHQVYKKEACIPCGSKNNODHSVCYSDCFYH 676
Db 603 TSCPAGYIIDRDSGCHSCPNTILKAHQYGVQACVPCGPGTKNNKIHSCLYNDCTFSR 662
Qy 677 EKENQILHYDFSNLSSVGLMNGPSTSGTKYFHFHFNISLCGHEGKKMALCTNNITDFT 736
Db 663 NPTPTFTFNYSALANTVTLAGGSPSTSGKLYFHFHFTLSLCGNOGRKMSVCTDNVTLR 722
Qy 737 VKEIVAGSDDYTNLVGAFVCSQSTIIPSESKGFRAALSSQSIILADTFVGVTTTLKNIN 796
Db 723 IPE---GESGFSKSIAYVCQAVIIPPEVTGYKAGVSSQPVSLADRLGTDTMLDGI 779
Qy 797 IKEDMFPVPTSQIPDVHFFYKSSATTTCINGRSTAVKMRNPTKSGAGVISVPSKCPAG 856
Db 780 SPALFHLESGLIPDVIFFYRNDVTQSCSSGRSTTIRVRCSPQKTVPGSLLLPCTCSDG 839
Qy 857 TDCGCTFYFLWESAAACPLCTEHDHFEIEGACKRGFQETLYVWNEPKWCIKISLPEKKL 916
Db 840 TDCGCFNFHFLWESAAACPLCSVADYHAIYSSCVAGIQKTTYVWREPKLCSGISLPEQRV 899
Qy 917 ATCEVDFWFLKVGAGVGAFTALLVALTCYFWKKKQKKK---KTLN 960
Db 900 TICKTIDFWLKVIGSAGCTAILLVLTLCYFWKKKQKLEYKYSKLVNM 947

RESULT 4
AAB35333
ID AAB35333 standard; Protein; 1001 AA.
XX AAB35333;
DT 08-MAY-2001 (first entry)
DE Human TR13 receptor protein SEQ ID NO: 40.
KW Human; tumour necrosis factor receptor; TR13; TR14; infection;
KW cancer; autoimmune disease; allergy; inflammatory disease;
KW graft rejection; apoptosis; cardiovascular disease; aneurysm.
OS Homo sapiens.
PN WO200105834-A1.
XX 25-JAN-2001.
XX 14-JUL-2000; 2000WO-US19343.
XX 16-JUL-1999; 99US-0144087.
PR 18-AUG-1999; 99US-0149450.
PR 20-AUG-1999; 99US-0149712.
PR 10-SEP-1999; 99US-0153089.
XX (HUMA-) HUMAN GENOME SCI INC.
PA Ruben SM, Ni J, Young PE;
XX WPI; 2001-112682/12.
XX Nucleic acids encoding 2 human tumor necrosis factor receptor
PT polypeptides (TR13) and (TR14), useful for the prevention, diagnosis
PT and treatment of, e.g. cancers, acquired immune deficiency syndrome and
PT hypohidrotic ectodermal dysplasia -
XX Claim 40; Page 398-401; 418pp; English.
PS The present invention provides the protein and coding sequences of the

CC human tumour necrosis factor receptors TR13 and TR14. These sequences are
CC useful in the diagnosis and treatment of many diseases, including cancer,
CC autoimmune diseases, cardiovascular disorders, allergies,
CC neurodegenerative diseases, graft rejection, inflammation; aneurysms and
CC infections.
XX Sequence 1001 AA;
SQ Query Match 52.6%; Score 2815.5; DB 22; Length 1001;
Best Local Similarity 52.8%; Pred. No. 1.1e-213;
Matches 501; Conservative 161; Mismatches 261; Indels 25; Gaps 12:
Qy 23 RGRSPWSPAMTCCHALAGCQANAG---DLPSSSRPLPPCOEKDYHFEYTECDSSGSR 79
Db 15 KQRTERRIPR---LWRL---LLWAGTAFQVQTGGPELHACKSESEYHFEYTACUSTSR 67
Qy 80 WRVALPNSAVDCSGLPDPVRKCECTFSCASGEYLEMKNOVCSCGEGTYSLSGSIKFDWM 139
Db 68 WRVAVPHTPGLCTSLPDPVKTECSFSCNAGEFLDMKQDSCKPCAEGRYSLSGTGIRFDEW 127
Qy 140 DELPAGFSNIAFTMDTVVGPDSRDPDGCNNNSWIPRGNTVIESNRDDCTVSLIYAVILKKS 199
Db 128 DELPHGFASLASANMELDDSAES-TGNCCTSSKWPVRGDIYAFTDECTATLMYAVNLKOS 186
Qy 200 GYVPEYQYVDNNIIFETFIQNDQCOEMDTTDDKVKLTDCGEMSGSHSVMLKSGTNILYW 259
Db 187 GYNFEIYYPDSIIIFEFVQNDQOP-NADSRWMTTEKG-WEFHSEVELNRGNVLYW 244
Qy 260 RTTGILMGSKAVKPVLVKNITIEGVAYTSECPKPGTFNKPQSGFNCQVCPNRYSEK 319
Db 245 RTAFSVWTKVPKPVLRNIAITGVAYTSECPKPGTGYADKQSGSFCKLPANSYSNKG 304
Qy 320 AKECIRCKDDQSFS--GSSECTERPPCTTKDYFQIHTPCDEBEGTKQIMYKNIETPKICRED 377
Db 305 ETSCHQC-DPDKYSEKSSSCNVRPACTDKDYFHTACDANGETQLMYKWKAPKICSED 363
Qy 378 LTDAIRLPSPGKKDCPCNPGFYNNGSSCHPCPGTSDGTKECPCPAGTEPALGFE 437
Db 364 LEGAVKLPSGKVTGHCPCNPGFFKTNSTCQPCPGVSYNNGS-DCTRCPAGTEPALGFE 422
Qy 438 YKWNVLPGNMKTSCFNVGNKCDGMNGVEAGDHIQSGAGGSDNDYLLNLHLPQFKPP 497
Db 423 YKWNVLPNTMETTVLSGINFEYKGMTGEVAGDHIYTAAGASDNDYLLNLVVPFRPP 482
Qy 498 TS-MTCATGSELGRITFVETELCSADCVLFWVDINRKTNNVSVESGCTKEQAYTHILF 556
Db 483 QSVMDTENKEVARITVFETLCSVNCLELYFMVGNSTNTPVETWKGSKQSYTYIE 542
Qy 557 KNATFTFAFORTNQODNRRFINDMWKYSITATNAVDGAVSSCRACALGSEQSGSSC 616
Db 543 ENTTSFTWAFORTTFHEASRYTNDVAKIYSINVTNVMNGVASCRCALASDVGSSC 602
Qy 617 VPCPPGHVIEKETNOCKECPDPTYLSIHQVYKKEACIPCGSKNNODHSVCYSDCFYH 676
Db 603 TSCPAGYIIDRDSGCHSCPNTILKAHQYGVQACVPCGPGTKNNKIHSCLYNDCTFSR 662
Qy 677 EKENQILHYDFSNLSSVGLMNGPSTSGTKYFHFHFNISLCGHEGKKMALCTNNITDFT 736
Db 663 NPTPTFTFNYSALANTVTLAGGSPSTSGKLYFHFHFTLSLCGNOGRKMSVCTDNVTLR 722
Qy 737 VKEIVAGSDDYTNLVGAFVCSQSTIIPSESKGFRAALSSQSIILADTFVGVTTTLKNIN 796
Db 723 IPE---GESGFSKSIAYVCQAVIIPPEVTGYKAGVSSQPVSLADRLGTDTMLDGI 779
Qy 797 IKEDMFPVPTSQIPDVHFFYKSSATTTCINGRSTAVKMRNPTKSGAGVISVPSKCPAG 856
Db 780 SPALFHLESGLIPDVIFFYRNDVTQSCSSGRSTTIRVRCSPQKTVPGSLLLPCTCSDG 839
Qy 857 TDCGCTFYFLWESAAACPLCTEHDHFEIEGACKRGFQETLYVWNEPKWCIKISLPEKKL 916
Db 840 TDCGCFNFHFLWESAAACPLCSVADYHAIYSSCVAGIQKTTYVWREPKLCSGISLPEQRV 899
Qy 917 ATCEVDFWFLKVGAGVGAFTALLVALTCYFWKKKQKKK---KTLN 960

breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha; adipocyte; ear; proliferation; glucose; free fatty acid; skeletal muscle; adipocyte; A-peptide; factor VIIA; gene therapy.

Homo sapiens.

WO200140466-A2.

07-JUN-2001.

01-DEC-2000; 2000WO-US32678.

01-DEC-1999; 99WO-US28301.

01-DEC-1999; 99WO-US28634.

02-DEC-1999; 99WO-US28551.

02-DEC-1999; 99WO-US28564.

02-DEC-1999; 99WO-US28565.

09-DEC-1999; 98US-0170262.

16-DEC-1999; 99WO-US30095.

20-DEC-1999; 99WO-US30911.

20-DEC-1999; 99WO-US30999.

30-DEC-1999; 99WO-US31243.

06-JAN-2000; 2000WO-US00277.

06-JAN-2000; 2000WO-US00376.

11-FEB-2000; 2000WO-US03565.

18-FEB-2000; 2000WO-US04341.

18-FEB-2000; 2000WO-US04342.

22-FEB-2000; 2000WO-US04414.

24-FEB-2000; 2000WO-US04914.

24-FEB-2000; 2000WO-US05004.

01-MAR-2000; 2000WO-US05601.

20-MAR-2000; 2000WO-US07377.

21-MAR-2000; 2000WO-US07532.

30-MAR-2000; 2000WO-US08439.

17-MAY-2000; 2000WO-US13705.

22-MAY-2000; 2000WO-US14042.

30-MAY-2000; 2000WO-US14941.

02-JUN-2000; 2000WO-US15264.

10-NOV-2000; 2000WO-US30873.

(GETH) GENENTECH INC.

WPI; 2001-408281/43.

N-PSDB; AAS21262.

Isolated, secretory and transmembrane PRO polypeptide used to detect other PRO polypeptides, link bioactive molecules to cells expressing PRO polypeptides, and detect the presence of mammalian tumours e.g. lung, breast, prostate, cervical.

Claim 12; Fig 38; 813pp; English.

AAU12172-AAU12446 represent novel human secretory and transmembrane PRO polypeptides. The PRO polypeptides are useful to detect other PRO polypeptides, to link bioactive molecules to cells expressing PRO polypeptides, to modulate biological activities of cells expressing PRO polypeptides, and to detect the presence of mammalian lung, colon, breast, prostate, rectal, cervical or liver tumours by comparing PRO polypeptide expression in a cell sample to that in a control sample. Some of the 275 sequences are also useful to stimulate the release of tumour necrosis factor-alpha (TNF-alpha) from human blood, the proliferation or differentiation of chondrocytes, the proliferation or gene expression in pericyte cells, the release of proteoglycans from cartilage, the proliferation of inner ear utricular supporting cells or of T-lymphocytes, the release of a cytokine from peripheral blood monocytes (PMBCs), or the proliferation of endothelial cells. Some of the PRO polypeptides may modulate glucose or free fatty acid uptake by skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide to factor VIIA. The PRO polypeptides can be used in assays to identify

molecules involved in binding interactions. The polynucleotides encoding PRO polypeptides can be used to generate probes, antisense RNA/DNA, transgenic or knock out animals and can be used in gene therapy.

SQ Sequence 1013 AA;

Query Match 52.2%; Score 2796; DB 22; Length 1013;
Best Local Similarity 53.4%; Pred. No. 3.9e-212;
Matches 494; Conservative 159; Mismatches 254; Indels 18; Gaps 10;

QY 46 WAG---DLPSSSRRLPPCQEKDYHFEYTECDSSSRWRVAIPNSAVDCSGLPDPVRKE 102

DB 31 MAGTAFQVTQGTGCPDLHACKSEYHYETACDSTGSRWRVAVPHTPGLCTSLDPVKTE 90

QY 103 CTFSCASGEYLEMKQVCSKCGEGTSLGSGIKFDEWDELPAFESNIAFTMDTVVGPSPS 162

DB 91 CSFSCNAGEFLDMKQSCCKPCAEGYSLGTGRFDEWDELPHGFPASLSANMELDDSAAS 150

QY 163 RPDGNNSSWIPRGNYIESNRDDCTVSLIYAVHLKSGYVFEYQYVDNIIFFEFIOND 222

DB 151 -TGNCSTSSKWPRGDYIASNTDECTATLMYAVNLKQSGTVNEEYYPDSIIPEFEVQND 209

QY 223 QCQEMDTTDDKWKVLTNDGCHWSHMLKSGTNILYWRITGILMGSKAVKPVLVKNITIE 282

DB 210 QCQP--NADDSRMKTKTEG--WEFHSELNRGNVLYWRTAFSVMTKVPKPLVLRNIAIT 267

QY 283 GVAYTSECFPCPKPTFSNKPQSGFNCQVCPNNTYSEKGAKECIRCKDDSOFS--GSSECTE 340

DB 268 GVAYTSECFPCPKPTYADKQSGSFCKLCPANSYSNKGESCHQC--DPDKYSEKSSSCNV 326

QY 341 RPPCTTKDYFQIHTPCDEBKTOIMYKWIETPKICREDLTDALRPPSGEKKDCCPNCNPF 400

DB 327 RPACTDKDYFYTHTACDANGETQLMYKWKPKICSEDLGAVKLPASGVKTHCPCPCNPF 386

QY 401 YNNGSSCHPCPGPTFSDGTKECRPCAGTEPALGFYKMWVLPGNMKTSCFNVGNSKC 460

DB 387 FKTNNSTCOPCYGYSNGS--DCTRCPAGTEPAVGFEYKMWNTLPNTMETTVLSGINFEY 445

QY 461 DGMNGWEVAGDHIQSGAGGSDNDYLILNLHIFGFKPPTS--MTGATGSELGRITFEFETLC 519

DB 446 KGMTGWEVAGDHIYTAAGASDNDFMLTLVVPGRPPQSVMAADTENKEVARITFEFETLC 505

QY 520 SADCVLVFWNDINRKSTNVVSWGGTKKQATVTHIIFKNATFTTFAFORTNOGQDNRRF 579

DB 506 SYNCELYFMVGNSTNTPEVTWKSGKQSYTYIEENTTTSTFWAFQRTTFFHEASRKY 565

QY 580 INDMVKIYSITATNAVDGVASSCRACALGSEQSGSCVPCPPGHYTEKETNOCKECPDPT 639

DB 566 TNDVAKIYSINTVNMVNGVASCRCALASDVGSCTSCPAGYIDRDSGTCHSCPPNT 625

QY 640 YLSIHQVYGEACIPCGPSKNNQDHSVCYDCFFYHEKENQILHYDFSNSLSSVGLMNG 699

DB 626 ILKAIHQVYGVQACVPCGPGTKNKKIHSCLYNDCTFSRNTPTRTFNYPNSALANTVTLAGG 685

QY 700 PSFTSKGYTFHFFNLSICGHEGKKMALCTNNITFTVKEIVAGSDDYTNLGVAFVQSTW 759

DB 686 PSFTSKGLKYFHFFHLSLGCNQGKMSVCTDNVTDLRPE---GESGFSKSIATYVQAV 742

QY 760 IIPSSKGFRAALSSQSITLADTFIGTVETTLKINIKEDMEPVPTSOIPDVHFFYKSS 819

DB 743 IIPPEVTGYKAGVSQPSVSLADRLIGVTDITDGLTSPAELFHLESIGIPDVFIFYSN 802

QY 820 TATTSCINGRSTAVKMRNCNPTKSGAGVISVPSKCPAGTCDCGCTFFFLMESAEACPLCTBH 879

DB 803 DVTQSCSSGRSTTIRVRCSPQKTPVPSGLLLPCTCSDGTCDCGNFHLWESAACPLCSVA 862

QY 880 DPHEIEGACKRGQETFLYVWNEPKWCIKISLPEKKLATCETVDFWLVKGVAGVGAFTAVL 939

DB 863 DYHAIIVSSCVAGIQXTTYVXREPKLCSGSIPEQRVTICKTIDFWLVKGVISAGTCTAIL 922

QY 940 LVALLTCYFWKKNNKKK---KTLIN 960

DB 923 LVVLTCYFWKKNNKKLEYKYSKLVMN 947

```

RESULT 7
AAB83850
ID AAB83850 standard; Protein; 911 AA.
XX
AC AAB83850;
XX
DT 23-JUL-2001 (first entry)
XX
DE Amino acid sequence of a human protein expressed in tumour cells.
XX
KW Tumour cell; immunological disease; autoimmune disease; cancer;
KW infection.
XX
OS Homo sapiens.
XX
PN WO200131003-A1.
XX
PD 03-MAY-2001.
XX
PF 30-OCT-2000; 2000WO-FR03032.
XX
PR 29-OCT-1999; 99FR-0013629.
XX
PA (FABR ) FABRE MEDICAMENT SA PIERRE.
XX
PI Delneste Y, Magistrelli G, Jeannin P, Bonnefoy J;
XX
DR WPI: 2001-328651/34.
XX
DR N-PSDB; AAF89774.
XX
PT New nucleic acid, expressed in tumours and lymphoid tissue is useful for
PT identifying agents for treating tumours and autoimmune disease
XX
PS Claim 10; Page 60-63; 85pp; French.
XX
CC The present sequence represents a human protein expressed in tumour
CC cells. The polynucleotide is useful for screening cDNA/genomic DNA banks
CC and for cloning isolated DNA; identifying mutant forms of the gene that
CC encodes a human protein, where the mutations are associated with
CC abnormal gene expression, or promoters and regulators of the gene,
CC particularly for diagnosis; for recombinant expression of the derived
CC protein; as probes and primers for detection and amplification; and
CC as antisense therapeutics. The tumour expressed protein is useful for
CC raising specific antibodies and to screen agents that modulate its
CC activity, bind to it or interact with it. These agents are potentially
CC useful for treatment or prevention of diseases associated with abnormal
CC expression/activity of the protein, particularly immunological diseases
CC (autoimmune diseases and cancer) or viral, bacterial, fungal or parasitic
CC infections.
XX
SQ Sequence 911 AA;

Query Match
Best Local Similarity 50.6%; Score 2712.5; DB 22; Length 911;
Matches 482; Conservative 156; Mismatches 253; Indels 21; Gaps 11;

QY 23 RGRSPWPSPAWTCWALAGCOAAWAG---DLPSSSSRPLPPCOEKDYHFEYTECDSSGSR 79
DQ 15 RGRTERRIIPR---LWRL---LLWAGTAFQVQTGTGCPHELHACKESHEYHVEYTACDSTGSR 67
QY 80 WRVATPNSAVDCGSLPDPVRGKECTFTSCASGEYLEMKNQVCKSCGEGTYSLSGSKGFDEW 139
DQ 68 WRVAVPHTPGLTSLPDPVKGTGCFSCNAGFELDMKDQSCPKCAEGRYSLGTGIRFDEW 127
QY 140 DELPAGESNIATFMTVVGPSDRPDGCGNNSWIPRGNIENRDCCTVSLIYAVHLKKS 199
DQ 128 DELPHGFASLSANMELDDSAES-TGNCSTSSKWVPRGDIASNTDECTATLMYAVNLKQS 186
QY 200 GYVFFEQYQVDNNEIFFEIONDQCEMDTTDKVKVLTNDGEGSHSVMLKSGTNILYW 259
DQ 187 GTVNEFYYPDSIIFFEYVQNDQCP-NADDSRWKMKTEG-WEFHVELNRGNVLYW 244

```

```

QY 260 RTTGILMGSKAVKPVLVKNITIEGVATSECPKPGCTFSNKGSCFNCQVCPNTYSEKG 319
DB 245 RTAFSVWTKVPKPVLRNIAITGVATSECFKPGCTGYADKOGSSFCFLCPANSYNGK 304
QY 320 AKECIRCKDDSOFS--GSSECTERPPCTTKDYFIHTPCDEEGKTQIMYKWIPEKICRED 377
DB 305 ETSQHOC-DPDKYSEKSSSCNVRPACTDKDYFYTHFACDANGETQLMYKWKAPKICSED 363
QY 378 LTDAILPLPSGKKKDCPCPNPGFYNNNGSSSCHPCPPGTFSDGTKECRPCPAGTEPALGFE 437
DB 364 LGAVALPASGVKTHCPNPGCFKTNNSTCQPCPYGSNGS-DCTRCPCAGTEPAVGFE 422
QY 438 YKWNVLPNCKMTSCFNVGNSCDGMNGEYAGDHTQSGAGGSDNDYLILNLHLPKPKPP 497
DB 423 YKWNVLPNCKMTSCFNVGNSCDGMNGEYAGDHTQSGAGGSDNDYLILNLHLPKPKPP 482
QY 498 TS-MTGATGSELGRITVFETILCSADCVLVFMVDINRSTNVVESMGTKKQAYTHIF 556
DB 483 QSVMDATENKEVARITVFETILCSADCVLVFMVDINRSTNVVESMGTKKQAYTHIF 542
QY 557 KNATFTFTWAFQRTNQGODNRFPINDMVKIYSITATNAVDGVASSCRACALGSEQSGSSC 616
DB 543 ENTFTSFTWAFQRTNQGODNRFPINDMVKIYSITATNAVDGVASSCRACALGSEQSGSSC 602
QY 617 VPCPPGHYIEKETNOCKECPDPTLYLSIHQVYGEACIPCGPGSKNODHSCVYSCDFFYH 676
DB 603 TSCPGAGYIDRDSGCHSCPPNTILKAHPYGVQACVCGPGTKNNKIHSLCYNDCFTSR 662
QY 677 EKENQILHYDFSNLMSGSLMNGSPSTSKGTYFHFHFNISLSCGHEGKMKALCTNNITDET 736
DB 663 NPTRTFNFNFSALANTVTLAGGSPSTSKGLYFHFHFTLSLSCGNQGRKMSVCTDNVTDLR 722
QY 737 VKEIVAGSDDYNLVGAFCVQSTIIPSESKGFRAALSSOIIILADTFIGVTVTTKLNIN 796
DB 723 IPE---GESGFSKSIYAYVQAVIIPPEVTGYKAGVSSQPVSLADRLIGVTTDMTDLGIT 779
QY 797 IKEDMPFVPTQIPDVHFFYKSTATTSCINGRSTAVKMRCPNPKSGAGVISVPKCPAG 856
DB 780 SPAELFHELSLGPDIIVFYRNDVYQSCSSGRSTIRVRCSPQKTVPGSLLLPGTCSOG 839
QY 857 TCDGCTFFYFLWSAEACPLCTEHDHFIEGACKRGQFETLYVMNEPKWCTIGISLPEKL 916
DB 840 TCDGCFNFHFLWSAEACPLCSVADYRAIVSSCVAGIQTKTYVWREPKLCSGGISLPEQRV 899
QY 917 ATCEVDFWLKV 928
DB 900 TTKTIDFWLKV 911

RESULT 8
AAB83851
ID AAB83851 standard; Protein; 870 AA.
XX
AC AAB83851;
XX
DT 23-JUL-2001 (first entry)
XX
DE Amino acid sequence of a human protein expressed in tumour cells.
XX
KW Tumour cell; immunological disease; autoimmune disease; cancer;
KW infection.
XX
OS Homo sapiens.
XX
PN WO200131003-A1.
XX
PD 03-MAY-2001.
XX
PF 30-OCT-2000; 2000WO-FR03032.
XX
PR 29-OCT-1999; 99FR-0013629.
XX

```

PA (FABR) FABRE MEDICAMENT SA PIERRE.
XX Delneste Y, Magistrelli G, Jeannin P, Bonnefoy J;
XX WPI; 2001-328651/34.
DR N-PSDB; AAF89775.
XX New nucleic acid, expressed in tumours and lymphoid tissue is useful for
PT identifying agents for treating tumours and autoimmune disease
XX
PS Claim 10; Page 67-70; 85pp; French.
XX
XX The present sequence represents a human protein expressed in tumour
CC cells. The polynucleotide is useful for screening cDNA/genomic DNA banks
CC and for cloning isolated DNA; identifying mutant forms of the gene that
CC encodes a human protein, where the mutations are associated with
CC abnormal gene expression, or promoters and regulators of the gene,
CC particularly for diagnosis; for recombinant expression of the derived
CC protein; as probes and primers for detection and amplification; and
CC as antisense therapeutics. The tumour expressed protein is useful for
CC raising specific antibodies and to screen agents that modulate its
CC activity, bind to it or interact with it. These agents are potentially
CC useful for treatment or prevention of diseases associated with abnormal
CC expression/activity of the protein, particularly immunological diseases
CC (autoimmune diseases and cancer) or viral, bacterial, fungal or parasitic
CC infections.
XX
SQ Sequence 870 AA;

Query Match 50.4%; Score 2702.5; DB 22; Length 870;
Best Local Similarity 54.1%; Pred. NO. 8e-205;
Matches 473; Conservative 153; Mismatches 237; Indels 11; Gaps 8;

QY 58 LPPCQEKDYHFEYTECDSSGSRWVAIPNSAVDCSLPDPVRGKECTFSCASGEYLEMKN 117
DB 1 LHACKSEHYEYACDSTGSRWVAIPHTPGCLTSLPDPVRGTECFSCNAGEFLDMKD 64

QY 118 QVCSKGEGETYSLGSIKDEWDELPAGFSNIATFMDTVVGPSPDCNNSNWTIPRGN 177
DB 65 QSCPKCAEGYSLGTGIRDEWDELPHGFASLSANMELDDSAES-TGNCSTSSKNVPRGD 123

QY 178 YTESNRDDCTVSLIYAHUKGSGYFFEQYVDNNIFEFFEQNDQCQEMDTTDDKWKVL 237
DB 124 YIASNTDECTATLMAVNLKQSGTVNFYDPSIIFFEQNDQCQ-PNADDSRWMTK 182

QY 238 TONGEGSHVMLKSTNTLYNRTTGILMGSKAVKPLVKNITIEGVAYTSECFCKPQGT 297
DB 183 TEGK-WEFHVELNRRNNVLYRTAFSWTKVPKPLVRLNIAITGVAYTSECFCKPQGT 241

QY 298 FSNKPGSFNCQVPRNTYSEKAKCIRCKDDSDQS--GSSECTERPPCTTKDYFQIHTP 355
DB 242 YADKQGSFCKLCPANSYKNKGETSCHQC-DPKYSEKSGSSCNVRPACTDRDYFHTA 300

QY 356 CDEEGTQIMYKWIPIKREDLTDAILRPPSGEKKDCPCPNPGFYNNNGSSSCHPCPPGT 415
DB 301 COANGETQLMYKWKPKICSEDLGAVKLPASGVKTHPCPNPGFPKFTNNSTQCPCPYGP 360

QY 416 FSDGTECRPCPAGTEPALGFYKKNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIOS 475
DB 361 YSNGS-DCTRCPAGTEPAVGFEYKWNLTLPNTNMTVLSGINFEYKGMTGWEVAGDHIYT 419

QY 476 GAGSDNDYLIINLHLPFGFKPPTS-MTGATGSELGRITVFETLCSADCVLPMVDINRK 534
DB 420 AAGASNDPMILLTVVGRFPQSVADTENKEVARITVFETLCSNCLFYMVGNSR 479

QY 535 STNVVESMGTEKQAYTHIFKNATFTTFAFQRTNQDQNRRTFNDVMYKYSITATNA 594
DB 480 TNPVETWKGSKQSYTYIENNTTSETFAFORTTFHEASRKYNDVAKIYSINVTW 539

QY 595 VQVASSCRACALGSQSSGSCVPCPPGHIYIEKETNOCKECPDPTVLSIHQVYKCACTP 654
DB 540 MNGVASYRCALASDVSSGTSCTGAGYIIDRSDGTSCHSCPNTTLKAHQPYGVQACVP 599

QY 655 CGPGSKNNQDHSVCYSDCFFYHEKENQILHYDFSNLSSVGSMLNMGSPSTSKCTKYEHFFN 714
DB 600 CGPGTKNNKIHLSCYNDCTFSRNTPTRTFNYSALANTVTLAGGSPSTSKGLKYFHHFT 659

QY 715 ISLCGHEGKKMALCTNNITDFTVKELVAGSDDYTNLVGAFVQOSTLIIPSESKGFRAALSS 774
DB 660 LSLCNOGRKMSVCVDNVVDLRIPE--GESGFSKSIAYVCOAVLIPPEVTGYKAGVSS 716

QY 775 QSIILADTFIGVTVETTLKINIKEDMPFVPTSQIPDVHFFYKSSATTTCTINGRSTAVK 834
DB 717 QPVSLADRLIGVTTDWTLDGITSAPLHLESLGIPDVIFFRSNDVTQSCSSGRSTTIR 776

QY 835 MRCNPTKSCAGVISVPSKCPACTGCTCTFYFLWESAECPLCTEHDHFIEGACRKGQJE 894
DB 777 VRCSQKTPVPSLLPGTCTGCTGDCGCFHFLWESAACPLCSVADYRAIVSSCVAGIGK 836

QY 895 TLYVWNEPKWCKIGSLPEKKLATCTETVDFWLKV 928
DB 837 TTYVMREPCLKSGGISLPEQRVTICKTIDFWMLKV 870

RESULT 9
AAB48377
ID AAB48377 standard; Protein: 464 AA.
XX
XX AAB48377;
XX
DT 20-APR-2001 (first entry)
XX
DE Human SEC10 protein sequence (clone ID 1795045-0.77).
XX
KW SECX: cytostatic; gynecological; gene therapy; screening assay; human;
KW SEC10: chromosomal mapping; forensic biology; cell proliferation; cancer;
KW cell differentiation; immune associated disorder; gestational disease.
XX
OS Homo sapiens.
XX
PN WO200078802-A2.
XX
PD 28-DEC-2000.
XX
PF 23-JUN-2000; 2000WO-US17328.
XX
PR 23-JUN-1999; 99US-0140584.
PR 20-JUL-1999; 99US-0144722.
PR 16-SEP-1999; 99US-0154520.
PR 22-JUN-2000; 2000US-0604286.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shinketsu RA, Fernandes E, Vernet C, Yang M, Boldog FL;
PI Hermann JL;
XX
DR WPI; 2001-071385/08.
DR N-PSDB; AAC84891.
XX
PT Polynucleotides encoding SECX proteins useful for treating disease
PT characterized by an aberrant level of cell proliferation and/or
PT differentiation like cancer or immune associated disorders
XX
PS Claim 1; Fig 10; 132pp; English.
XX
XX The invention relates to human SECX polypeptides and polynucleotides
CC encoding them. The SECX polypeptides can be expressed by standard
CC recombinant methodology. The SECX polypeptides are useful for treating
CC or preventing a SECX-associated disorder. The invention is useful in
CC screening assays; detection assays (e.g. chromosomal mapping, cell and
CC tissue typing, forensic biology); predictive medicine (diagnostic assays,
CC prognostic assays, monitoring clinical trials, and pharmacogenomics); and
CC methods of treatment (e.g. therapeutic and prophylactic), especially
CC disorders characterized by aberrant cell proliferation and/or
CC differentiation like cancer or immune associated disorders or gestational
CC disease. The present sequence represents a SEC10 protein.

```
XX SQ Sequence 464 AA;
Query Match 47.2%; Score 2526; DB 22; Length 464;
Best Local Similarity 98.9%; Pred. No. 3.1e-191;
Matches 452; Conservative 0; Mismatches 3; Indels 2; Gaps 1;

QY 115 MKNQVCSKCGECTYSLGSIKFDEWDELPAGFSNIATFMDTVVGPDSRPGDGNSSWIP 174
DB 1 MKNQVCSKCGECTYSLGSIKFDEWDELPAGFSNIATFMDTVVGPDSRPGDGNSSWIP 60
QY 175 RGNYESNRDDCTVSLIYAVHLKKSQYVFFQYVYVDDNIFFEFFQNDQCOEMDTTDDKW 234
DB 61 RGNYESNRDDCTVSLIYAVHLKKSQYVFFQYVYVDDNIFFEFFQNDQCOEMDTTDDKW 120
QY 235 VKLTONGEGSHVSMKSGTNILYWRITGILMGSKAVKPVLVKNITIGVAYTSECFFPK 294
DB 121 VKLTONGEGSHVSMKSGTNILYWRITGILMGSKAVKPVLVKNITIGVAYTSECFFPK 180
QY 295 PGTFSNKGSGFNCQVCPNRYSEKAKECIRCKDDSQFS--GSSSECTERPPCTTKDYFOI 352
DB 181 PGTFSNKGSGFNCQVCPNRYSEKAKECIRCKDDSQFSSEGSSECTERPPCTTKDYFOI 240
QY 353 HTPCDEEGKTQIMYKWIIEPKICREDLTDAILRPPSGEKKDCPPCNPFGYNNSSSCHPCP 412
DB 241 HTPCDEEGKTQIMYKWIIEPKICREDLTDAILRPPSGEKKDCPPCNPFGYNNSSSCHPCP 300
QY 413 PGTFSNKGSGFNCQVCPNRYSEKAKECIRCKDDSQFS--GSSSECTERPPCTTKDYFOI 472
DB 301 PGTFSNKGSGFNCQVCPNRYSEKAKECIRCKDDSQFS--GSSSECTERPPCTTKDYFOI 360
QY 473 IQSGAGSGDNDYLILNLHIPGKPTMTGATGSELGRITTFVETLCSADCVLYFWVDIN 532
DB 361 IQSGAGSGDNDYLILNLHIPGKPTMTGATGSELGRITTFVETLCSADCVLYFWVDIN 420
QY 533 RKSTNVVESWGGTKEQAVTHIIFKNATFTTWAQF 569
DB 421 RKSTNVVESWGGTKEQAVTHIIFKNATFTTWAQF 457

RESULT 10
AAB48372
ID AAB48372 standard; Protein; 411 AA.
AC AAB48372;
XX 20-APR-2001 (first entry)
XX Human SEC5 protein sequence (clone ID 1795045.0.61).
XX SEC5; cytostatic; gynecological; gene therapy; screening assay; human;
KW SEC5; chromosomal mapping; forensic biology; cell proliferation; cancer;
KW cell differentiation; immune associated disorder; gestational disease.
XX Homo sapiens.
OS WO200078802-A2.
PN 28-DEC-2000.
XX 23-JUN-2000; 2000WO-0517328.
XX 23-JUN-1999; 99US-0140584.
PR 20-JUL-1999; 99US-0144722.
PR 16-SEP-1999; 99US-0154520.
PR 22-JUN-2000; 2000US-0604286.
XX (CURA-) CURAGEN CORP.
XX Shimkets RA, Fernandes E, Vernet C, Yang M, Boldog FL;
PI Herrmann JL;
XX WPI; 2001-071385/08.
```

```
DR N-PSDB; AAC84886.
XX Polynucleotides encoding SEC5 proteins useful for treating disease
PT characterized by an aberrant level of cell proliferation and/or
PT differentiation like cancer or immune associated disorders -
XX Claim 1; Flg 6; 132pp; English.
PS The invention relates to human SEC5 polypeptides and polynucleotides
CC encoding them. The SEC5 polypeptides can be expressed by standard
CC recombinant methodology. The SEC5 polypeptides are useful for treating
CC or preventing a SEC5-associated disorder. The invention is useful in
CC screening assays; detection assays (e.g. chromosomal mapping, cell and
CC tissue typing, forensic biology); predictive medicine (diagnostic assays,
CC prognostic assays, monitoring clinical trials, and pharmacogenomics); and
CC methods of treatment (e.g. therapeutic and prophylactic), especially
CC disorders characterized by aberrant cell proliferation and/or
CC differentiation like cancer or immune associated disorders or gestational
CC disease. The present sequence represents a SEC5 protein.
XX SQ Sequence 411 AA;
Query Match 41.5%; Score 2222; DB 22; Length 411;
Best Local Similarity 98.5%; Pred. No. 2.9e-167;
Matches 397; Conservative 1; Mismatches 3; Indels 2; Gaps 1;

QY 169 NSSWTPRGNYIESNRDDCTVSLIYAVHLKKSQYVFFQYVYVDDNIFFEFFQNDQCOEMD 228
DB 2 HSSWTPRGNYIESNRDDCTVSLIYAVHLKKSQYVFFQYVYVDDNIFFEFFQNDQCOEMD 61
QY 229 TTTDKWKLTDONGEGSHVSMKSGTNILYWRITGILMGSKAVKPVLVKNITIGVAYTS 288
DB 62 TTTDKWKLTDONGEGSHVSMKSGTNILYWRITGILMGSKAVKPVLVKNITIGVAYTS 121
QY 289 ECFPCPKGTFSNKGSGFNCQVCPNRYSEKAKECIRCKDDSQFS--GSSSECTERPPCTT 346
DB 122 ECFPCPKGTFSNKGSGFNCQVCPNRYSEKAKECIRCKDDSQFSSEGSSECTERPPCTT 181
QY 347 KDYFOIHTPCDEEGKTQIMYKWIIEPKICREDLTDAILRPPSGEKKDCPPCNPFGYNNSS 406
DB 182 KDYFOIHTPCDEEGKTQIMYKWIIEPKICREDLTDAILRPPSGEKKDCPPCNPFGYNNSS 241
QY 407 SCHPCPPGTFSNKGSGFNCQVCPNRYSEKAKECIRCKDDSQFS--GSSSECTERPPCTT 466
DB 242 SCHPCPPGTFSNKGSGFNCQVCPNRYSEKAKECIRCKDDSQFSSEGSSECTERPPCTT 301
QY 467 EVAGDHIOGAGSGDNDYLILNLHIPGKPTMTGATGSELGRITTFVETLCSADCVLY 526
DB 302 EVAGDHIOGAGSGDNDYLILNLHIPGKPTMTGATGSELGRITTFVETLCSADCVLY 361
QY 527 FMVDINRKSTNVVESWGGTKEQAVTHIIFKNATFTTWAQF 569
DB 362 FMVDINRKSTNVVESWGGTKEQAVTHIIFKNATFTTWAQF 404

RESULT 11
AAB35328
ID AAB35328 standard; Protein; 750 AA.
XX AC AAB35328;
XX 08-MAY-2001 (first entry)
XX Human TR13 receptor protein SEQ ID NO: 2.
XX Human; tumour necrosis factor receptor; TR13; TR14; infection;
KW cancer; autoimmune disease; allergy; inflammatory disease;
KW graft rejection; apoptosis; cardiovascular disease; aneurysm.
XX Homo sapiens.
OS WO200105834-A1.
PN
XX
```

Db 577 GSLLLPTGTCSDGTCDGCMHFELWESAACPLCSVADYHAIVSSCVAGIOKTTTVWREPKL 636
 QY 905 CIKSGISPEKKLATCETVDFWLKVGAGVGAFPTALLVLTCTYFWKKNQKKK----KTILN 960
 Db 637 CSGGISLPEQVRVTKTIDFWLKVGISAGTCTALLTLVLTCTYFWKKNQKLEYKYSKLVMN 696

RESULT 12
 AAY59972
 ID AAY59972 standard; Protein; 495 AA.
 XX
 AC AAY59972;
 XX
 DT 31-JAN-2000 (first entry)
 XX
 DE Human endometrium tumour EST encoded protein 32.
 XX
 KW Endometrium; human; tumour; cancer; anticancer; cytostatic; EST;
 KW treatment; uterine; gene therapy; expressed sequence tag.
 XX
 OS Homo sapiens.
 XX
 PN DEL9817948-A1.
 XX
 PD 21-OCT-1999.
 XX
 PF 17-APR-1998; 98DE-1017948.
 XX
 PR 17-APR-1998; 98DE-1017948.
 XX
 PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
 XX
 PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E
 XX WPI: 1999-591957/51.
 DR N-PSDB: AAZ41991.
 XX
 PT New nucleic acid sequences expressed in uterine cancer tissues, and
 PT derived polypeptides, for treatment of uterine and endometrial cancer
 PT and identification of therapeutic agents .
 XX
 PS Claim 23; Page 287; 444pp; German.
 XX
 CC This invention describes novel human nucleic acid (cDNA) sequences (A)
 CC that are highly expressed in uterine tumour tissue and which have
 CC anticancer and cytostatic activity. (A) are used (i) for recombinant
 CC expression of polypeptides (B) and (ii) to isolate complete genes. (B)
 CC are used (i) to identify agents suitable for treatment of uterine or
 CC endometrial cancer; (ii) directly for treating these forms of cancer
 CC (including expression from gene therapy vectors) and (iii) for
 CC generation of specific antibodies. (A) are identified by assembling ESTs
 CC (expressed sequence tags) from a particular tissue type before compar-
 CC ing expression patterns. This allows a significantly longer fragment of
 CC the gene to be revealed, so should reduce the number of failures
 CC associated with the fact that ESTs from different libraries may represent
 CC different parts of the same unknown gene, distorting the estimated
 CC frequency of occurrence in a particular tissue. AAY59941-Y60328 represents
 CC protein fragments encoded by the human endometrium tumour cDNA library
 CC derived EST fragments represented in AAZ41981-Z42121.
 XX
 SQ Sequence 495 AA;

Query Match 23.5%; Score 1258.5; DB 20; Length 495;
 Best Local Similarity 50.8%; Pred. No. 6.8e-91;
 Matches 219; Conservative 81; Mismatches 124; Indels 7; Gaps

Db 62 VMNGVASYCRCALEASDVGSSTSCPAGYIDRDSCTCHSCPNTILKAHQPYGOACV 121
QY 634 PCGPGSKNNQDHSVCYDCFFYHEKENOILHYDFSNSLSSVGLMNGSPFTSGTKYKTHFF 713
Db 122 PCGPGTKNNKIHSLCYNDCTFSRNTPTRTFNYSALANTVTLAGGSPFTSKGLKYFHF 181
QY 714 NISLCHGSEKKWALCTNITDFTYKEIVAGSDDYTNLVGAFVCSOTIIPSESKGFRAALS 773
Db 182 TUSLCGNGRKMVSCTDNVTDLRPE---GESGFSKSAITAYCOAVIIPPEVTGKAGVS 238
QY 774 SOSIILATFIGVIVETTLKINIKEDMPVPTSQIPDVHFFYKSSSTATTCINGRSTAV 833
Db 239 SQPVSLADRLIGVTTDMTLDGITSAPALFHLIESLIGIPDVIFFYRSNDVTQSCSSGRSTTI 298
QY 834 KMRCPNTSGAGVISVPKCPAGTCOGCTFFFLWESAACPLCTEHDHEGACKRGFO 893
Db 299 RYRCSPOKTPVPSGLLPGTCDGCGNFHFLWESAACPLCSVADYHAIYSSCVAGIQ 358
QY 894 ETLVYVNEPKWCIKIGISLPEKKLATCETVDFWLKYGAGVGAFTAVLLVALTCYFWKKNQK 953
Db 359 KTYVWREPKLCSGGISLPEQRTICKTIDFWLKVGISAGTCTAILLVLCYFWKKNQK 418
QY 954 KK----KTILN 960
Db 419 LEYKYSKLVMN 429
RESULT 13
ID AAB83853
XX AAB83853 standard; Protein; 383 AA.
AC AAB83853;
XX
DT 23-JUL-2001 (first entry)
XX
DE Amino acid sequence of a human protein expressed in tumour cells.
DE
DE Tumour cell; immunological disease; autoimmune disease; cancer;
KW infection.
KW
XX Homo sapiens.
OS
XX
XX WO200131003-A1.
XX
XX 03-MAY-2001.
XX
PF 30-OCT-2000; 2000WO-FR03032.
XX
XX 29-OCT-1999; 99FR-0013629.
XX
XX (FABR) FABRE MEDICAMENT SA PIERRE.
PA
XX
XX Delneste Y, Magistrelli G, Jeannin P, Bonnefoy J;
PI
XX WPI; 2001-328651/34.
DR
DR N-PSDB; AAF89777.
XX
XX
PT New nucleic acid, expressed in tumours and lymphoid tissue is useful for
PT identifying agents for treating tumours and autoimmune disease
XX
XX Claim 10; Page 74-75; 85pp; French.
XX
XX The present sequence represents a human protein expressed in tumour
CC cells. The polynucleotide is useful for screening cDNA/genomic DNA banks
CC and for cloning isolated DNA; identifying mutant forms of the gene that
CC encodes a human protein, where the mutations are associated with
CC abnormal gene expression, or promoters and regulators of the gene,
CC particularly for diagnosis; for recombinant expression of the derived
CC protein; as probes and primers for detection and amplification; and
CC as antisense therapeutics. The tumour expressed protein is useful for
CC raising specific antibodies and to screen agents that modulate its
CC activity, bind to it or interact with it. These agents are potentially
CC useful for treatment or prevention of diseases associated with abnormal

CC expression/activity of the protein, particularly immunological diseases
CC (autoimmune diseases and cancer) or viral, bacterial, fungal or parasitic
CC infections.
XX
SQ Sequence 383 AA;
Query Match 22.6%; Score 1208.5; DB 22; Length 383;
Best Local Similarity 58.3%; Pred. No. 4.3e-87;
Matches 215; Conservative 53; Mismatches 94; Indels 7; Gaps 6;
QY 167 CUNSWITPRNGYIESNRDDCTVSLIYAVHLKKSIVFEYQYVDNNIFFEFFIQQDCQE 226
Db 19 CTSSKWPGRGDYIASNDECTATLMYAVNLKSGTVNFYYPDSSIIFFEFFVQDCQP 78
QY 227 MDTTDDKWKLTNGEGSHVLMKSGTNILYWRTTGILMGSKAVKPVLVKNITIEGVAY 286
Db 79 -NADDSRMKTKTEKG-WEFHSVELNRGNVLYWTTAFSWTKYKPVLYVRNIALTGVAY 136
QY 287 TSECFPCPKPGTFSNKPQSFNCQVCPRNTYSEKGAKEICRCKDDSQFS--GSSECTERPPC 344
Db 137 TSECFPCPKPGTYADKQSSFCCKLCPANSYSNKGTSCHQC-DPDKYSEK3SSSCNVRPAC 195
QY 345 TTQDYFQIHTPCDEBEGKTOIMYKWIPEKICREDLTDALRPPSGEKKDCPCNPGFYNNG 404
Db 196 TDKDYFYTHTTACDANGETQLMYKWKAKPKICSEDLGAVKLPAASGVKTHCPCNPGFFXTN 255
QY 405 SSSCHPCPPGTFSDGTEKRCPCPAGTEPALGFYKWNVLPGNMKTSCFNVNGSKCQDMN 464
Db 256 NSTCPCPGYPSNGS-DCTRCPAGTEPAVGFEYKWNVLTPTNMTETVLSGINFYKGMT 314
QY 465 GWEVAGDHIQSGAGSDNDYLILNLHIFPGKPPTS-MTGATGSELGRITFFETLCSADC 523
Db 315 GWEVAGDHIYTAAGASDNDFMILTIVPGFRPPQSMADTENKEVARITFFETLCSVNC 374
QY 524 VLYFNVMDIN 532
Db 375 ELYFNVGVN 383
RESULT 14
AAB85768
ID AAB85768 standard; Protein; 372 AA.
XX
AC AAB85768;
XX
DT 29-OCT-2001 (first entry)
XX
DE Human seven-transmembrane protein 50288 sequence.
XX
KW seven-transmembrane protein; G-protein coupled receptor; GPCR; human;
KW 17724; 50288; 31945; antiinflammatory; antitumor; cytostatic; virucide;
KW hepatotropic; immunosuppressive; gynecological; neuroprotective;
KW anti-HIV; immunostimulant; dermatological; antithyroid; cardiatic;
KW antianemic; antiParkinsonian; nephrotropic; antithyroid; hemostatic;
KW cerebroprotective; osteopathic; analgesic; gene therapy; nootropic.
OS
XX Homo sapiens.
XX
XX WO200159117-A2.
XX
XX 16-AUG-2001.
XX
PF 12-FEB-2001; 2001WO-US04536.
XX
XX 11-FEB-2000; 2000US-0182061.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Glucksmann MA, Silos-Santiago I;
PI
XX WPI; 2001-514670/56.
DR
DR N-PSDB; AAH76195, AAH76196.
XX

PT New seven-transmembrane protein/G-protein coupled receptor polypeptides
PT and polynucleotides for diagnosing, treating seven-transmembrane
PT protein/receptor-related disorders and to identify modulators of
XX therapeutic use

PS Claim 8; Page 139-141; 144pp; English.

XX The invention provides isolated seven-transmembrane protein/G-protein
CC coupled receptor polypeptides selected from 17724, 50288, 31945 proteins.
CC The polypeptides can be expressed by standard recombinant methodology.
CC Modulators of the polypeptides can be identified using a competition
CC binding assay or an assay for receptor-mediated signal transduction. The
CC polypeptides and polynucleotides are useful as reagents or targets in
CC seven-transmembrane protein/receptor assays applicable to treatment and
CC diagnosis of seven-transmembrane protein/receptor-mediated disorders
CC (see AAH76191 for a detailed description of the various disorders that
CC can be treated or diagnosed using the polypeptides). The polynucleotides
CC are useful to detect mutations in genes and gene expression products such
CC as mRNA, as antisense constructs to control gene expression and for
CC chromosome identification. The present sequence represents the human
XX seven transmembrane protein 50288 sequence.

XX Sequence 372 AA;

Query Match 16.5%; Score 883; DB 22; Length 372;
Best Local Similarity 51.7%; Pred. No. 2.4e-61;

Matches 163; Conservative 54; Mismatches 84; Indels 14; Gaps 7;

QY 23 RGRSPWPAWICWALAGCAQAAWAG---DLPSSSSRPLPPCKEYVHRYETCDSSGSR 79

DB 15 RGRTERIPR---LWRL-----LLWAGTAFQVTOGTGPELHACKSEYHYETACDSTGSR 67

QY 80 WRVAIPNSAVDCGLDPVVGKECTSCASGEYLEMKNQVCSCGEGTYSLSGIRFDEW 139

DB 68 WRVAIPHTPGCLTSLDPVVGKTECSFSCNAGEFLDMKQSCKPCAGRYSLGTGIRFDEW 127

QY 140 DELPAGFSNIATFMDTVVGPSDRPGCCNNSWIPRGNIETSRDQCTVSLIYAVILKKS 199

DB 128 DELPHGFASLSANWELDDSAES--TGNCSTSSKWVPGDYIASNTDECTATLYAVNLKOS 186

QY 200 GYVFFEQYVNDNIFFETQNDQCEMDTTTDKVKLTNDGWSHSLKSGTNILYW 259

DB 187 GTVNFYYPDSSITIEFFVQNDQCP--NADDSRWKMTTEKG--WEPHSVELNRGNVLYW 244

QY 260 RTTGILMGSKAVKPLVKNITIEGVAYTSECPCPKGTSTNKPSTNCOVCPNTYSEKG 319

DB 245 RTTAFSWTKVPKPLVLRNITAGVAYTSECPCPKGTADKQGSFCKLCPANSYSNKG 304

QY 320 AKECIRCKDDSQFSG 334

DB 305 ETSCHQC-DPDKYSG 318

RESULT 15

AAB83852

ID AAB83852 standard; Protein; 209 AA.

XX AAB83852;

XX 23-JUL-2001 (first entry)

XX Amino acid sequence of a human protein expressed in tumour cells.

XX Tumour cell; immunological disease; autoimmune disease; cancer;

XX infection.

XX Homo sapiens.

XX WO200131003-A1.

XX 03-MAY-2001.

XX 30-OCT-2000; 2000WO-FR03032.

XX 29-OCT-1999; 99FR-0013629.

XX (FABR) FABRE MEDICAMENT SA PIERRE.

XX Delneste Y, Magistrelli G, Jeannin P, Bonnefoy J;

XX WPI; 2001-328651/34.

XX N-PSDB; AAF89776.

XX New nucleic acid, expressed in tumours and lymphoid tissue is useful for
XX identifying agents for treating tumours and autoimmune disease

XX Claim 10; Page 71-72; 85pp; French.

XX The present sequence represents a human protein expressed in tumour
CC cells. The polynucleotide is useful for screening cDNA/genomic DNA banks
CC and for cloning isolated DNA; identifying mutant forms of the gene that
CC encodes a human protein, where the mutations are associated with
CC abnormal gene expression, or promoters and regulators of the gene,
CC particularly for diagnosis; for recombinant expression of the gene,
CC protein; as probes and primers for detection and amplification; and
CC as antisense therapeutics. The tumour expressed protein is useful for
CC raising specific antibodies and to screen agents that modulate its
CC activity, bind to it or interact with it. These agents are potentially
CC useful for treatment or prevention of diseases associated with abnormal
CC expression/activity of the protein, particularly immunological diseases
CC (autoimmune diseases and cancer) or viral, bacterial, fungal or parasitic
CC infections.

XX Sequence 209 AA;

Query Match 13.2%; Score 705.5; DB 22; Length 209;
Best Local Similarity 56.2%; Pred. No. 1.2e-47;

Matches 126; Conservative 26; Mismatches 55; Indels 17; Gaps 3;

QY 310 CPRNTYSEKAKECIRCKDDSQFSGSECTERPPCTTKDYFOIHTPCDEEKGTOIMYKWI 369

DB 2 CDPKYSEK-----GSSSCNVRPACTDKDYFTHTACDANGETQLMYKWA 46

QY 370 EPKICREDLTDAIRLPPSGEKKDCPPCPGYNNGSSSCHPCPPGTFSDGTYKECRCPAG 429

DB 47 KPKICSEDLGAVKLPASGVKTHCPNPGFETKNTNSTCQPCPYGPNYSNGS-DCTRCPCAG 105

QY 430 TEPALGCFEYKWNVLPPGNKTSCEFNVGNSKCDGMNGWEVAGDHQSGAGSDNDYLLNL 489

DB 106 TEPAVGFEYKWNVTLPNNKTTVLGSGINFEYKGTGWEVAGDHIYTAAGASDNDYLLNL 165

QY 490 HIPGFKPPTS-WTGATGSELGRITFVFETLCSADCVLYFMVDIN 532

DB 166 VVPGFRPPQSVNADPENKEVARITFVFETLCSVNCLEYFMVGVN 209

Search completed: May 12, 2003, 13:18:51

Job time : 51.4563 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 13:07:53 ; Search time 20.8085 Seconds
(without alignments)
1361.665 Million cell updates/sec

Title: US-10-073-333A-2
Perfect score: 5357
Sequence: 1 MLFRAGPVRGRGWGRPAEA.....TCYFWKKNQKKKTLNLFN 963

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	164	3.1	3111	2	US-08-460-309-4
2	164	3.1	3111	2	US-08-125-077-4
3	162	3.0	1253	3	US-08-479-722B-4
4	155.5	2.9	1251	5	PCT-US95-02251-3
5	155.5	2.9	1252	1	US-08-199-780-3
6	155.5	2.9	1252	2	US-08-316-650-3
7	151	2.8	2594	4	US-08-718-388-7
8	151	2.8	5405	4	US-08-718-388-9
9	149	2.8	610	1	US-08-365-470-3
10	149	2.8	610	3	US-09-209-668-19
11	148	2.8	610	4	US-09-009-490A-89
12	148	2.8	2703	1	US-08-185-432-19
13	148	2.8	2703	4	US-08-899-232-4
14	145	2.7	610	6	5217870-2
15	145	2.7	1193	2	US-08-400-159-10
16	145	2.7	1193	3	US-08-611-729A-10
17	143.5	2.7	2523	3	US-08-185-432-18
18	143.5	2.7	2523	4	US-08-899-232-3
19	142.5	2.7	1940	2	US-08-644-271-30
20	142.5	2.7	1940	4	US-09-077-955-34
21	139	2.6	1219	4	US-08-882-046-5
22	137.5	2.6	833	4	US-09-013-895A-5
23	137.5	2.6	833	4	US-09-448-868-5
24	137	2.6	1010	4	US-08-882-046-7
25	137	2.6	1036	4	US-09-068-740A-6
26	137	2.6	1187	4	US-09-088-740A-7
27	137	2.6	1208	4	US-09-199-865-1

28	137	2.6	1218	2	US-08-400-159-6
29	137	2.6	1218	3	US-08-611-729A-6
30	137	2.6	1218	4	US-08-882-046-2
31	137	2.6	1218	4	US-09-214-278-7
32	137	2.6	1218	4	US-09-068-740A-11
33	137	2.6	1964	4	US-09-467-997-1
34	136.5	2.5	2471	1	US-08-185-432-16
35	136.5	2.5	2471	1	US-08-083-590A-19
36	136.5	2.5	2471	3	US-08-532-384-19
37	136.5	2.5	2471	4	US-08-899-232-1
38	135.5	2.5	810	2	US-08-820-170A-34
39	135.5	2.5	810	3	US-09-055-699-34
40	135.5	2.5	810	4	US-09-273-565-34
41	135.5	2.5	810	4	US-09-565-538-34
42	135.5	2.5	810	4	US-09-661-468-34
43	133.5	2.5	299	4	US-09-286-529-17
44	133	2.5	300	2	US-08-794-796-2
45	132.5	2.5	1198	4	US-09-245-041-131

ALIGNMENTS

RESULT 1
US-08-460-309-4
; Sequence 4, Application US/08460309
; Patent No. 5837496
; GENERAL INFORMATION:
; APPLICANT: Engvall, Eva
; APPLICANT: Leivo, Ilmo
; TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
; TITLE OF INVENTION: Fragments and Uses Thereof
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460.309
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/125,077
; FILING DATE: 22-SEP-1993
; APPLICATION NUMBER: US PCT/US 94/10730
; FILING DATE: 21-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/472,319
; FILING DATE: 30-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/919,951
; FILING DATE: 27-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LA 9721
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3111 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-460-309-4

Db 703 VNLESAVSYPDTGSIATAAVEVCQPPGYTGSSCESCWPRHRRVNGTIFGGICEPCQCFGH 762
Qy 624 --YIEKETNOCKEPPDTYLSIHQVYKEACIPCPG-----SKNODHSVCYSDCCFF 675
Db 763 AESCDDVTGECLNCKDHT-----GGPYCDKCLPFGYGETGTGSDCQPCACPLNIP 814
Qy 676 HEKENQILHDFNSLSSVGLMNGPSETSKGTKYFHFNISLC--GHEGKKMALCT----- 729
Db 815 SNFSPCHLD-----RSLGLICDG-----CPVGYTGPCCRCABGYF 852
Qy 730 -----NNITDFTVEIVAGSDDYTNLGA-FVOCSTIIPSESKGFRAALS 773
Db 853 GQSPVPGSGOPOCNDNLDFS-----IPGSCD--SLSGCLICRP-----GTTG 895
Qy 774 SOSIILADFTIGVAVETTLKNINIKEDMFVPVTSQIPDVHFFYKSSNATTTSCINGRSTAV 833
Db 896 RYCELADGYGDAVDA--KNCQ-----PCRCNAGGSFSEVCHSOTGQCEC-----RANVQ 944
Qy 834 KMRNPTKSGAGVISVPSKCPAGTCDGCTFYFLWESAEACPLCTEHDFHEIEGACKRGFQ 893
Db 945 QORCD-----KCKAGT-----FGLQARGCVPCNCNSFGSKSPDCESGQ 984
Qy 894 ETLVWNEPKWCIKGISLPEKKLATC 919
Db 985 -----CMCQPGVT--GKKCDRC 999

RESULT 3

US-08-479-722B-4
; Sequence 4, Application US/08479722B
; Patent NO. 6074840
; GENERAL INFORMATION:
; APPLICANT: Bonadio, Jeffrey
; APPLICANT: Yin, Wushan
; TITLE OF INVENTION: LATENT TGF(BINDING PROTEIN (LBP)
; TITLE OF INVENTION: GENES, COMPOSITIONS AND METHODS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Williams, Morgan & Amerson
; STREET: 7676 Hillmont, Suite 250
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77040
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,722B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US95/02251
; FILING DATE: 21-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/316,650
; FILING DATE: 30-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/199,780
; FILING DATE: 18-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fussey, Shelley P.M.
; REGISTRATION NUMBER: 39,458
; REFERENCE/DOCKET NUMBER: 4100.000500/FUS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (713) 934-7000
; TELEFAX: (713) 934-7011
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1253 amino acids
; TYPE: amino acid

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-479-722B-4

Query Match 3.0%; Score 162; DB 3; Length 1253;
Best Local Similarity 20.3%; Pred. No. 1.4e-05;
Matches 150; Conservative 67; Mismatches 246; Indels .276; Gaps 48;

Qy 289 ECFCKP--GTFESKPKPSFNCQVCP--NTYSEKGAKECIRCKDDSQSGSSECTERPPC 344
Db 601 EAPCPGKGGKTCMTGSSYNCH--CNRGYRLHVAGGRSCV-----DLNECAKPHLC 650
Qy 345 TTKDY---FQIHTPCDEBKQIMV-----KWIEPKICREDLTDAIRLP---PSCEKDC 393
Db 651 GDGFCINFPHYKCN-----CYPGYRLKASRPIC-EDI-DECRDPSTCPDCKCNK 701
Qy 394 PCNPFGYNNSSSCHPCPPGTFSDGTKECR---PCPAGTEPALGFYKWNVLPGNMKT 450
Db 702 P-----GSFKCIACQPGYRSGGGACRDVNECSEGTPCSPG---WCENLPGSYRC 748
Qy 451 SC-----FNVNSKC---DGMNGWEVAGDHIOAGGSDNDYLIUNLHIPGFKPPTSMTG 502
Db 749 TCAQGIKTRTGLRLCIDVDECEAGKVCQDGTCTNTPGFCQCLSGYHL----- 797
Qy 503 ATGSELGRITFV---FETLC-SADCVLYFMVDINRKSTNVVESW-----GGTK 546
Db 798 --SRDRSCEIDIECDPPAACIGGDCI-----NTNGSYRCLCPLGHLVGGRK 843
Qy 547 -----EKQAYTHIIFKNATFTT---WAFQRTNOGQDNRR-FIN- 581
Db 844 CKKDIDECSDQPLCLPHACENLQGSYVCVDEGFTLTQDQHGCEEVEQPHHKKECYLNF 903
Qy 582 -DMVKIYSITATNAVDGVASSCRACALGSEQSGSC---VPC-----PPGHVIE 626
Db 904 DDTVFCDSVLATNVTQ---QEC-CCSLGAGW-GDHCEIYPCPVYSSAEFHSLLVDPDKRLH 958
Qy 627 KETNOCKECPDPTLYLSIHQ---VYKKEACI---PCGPGSKNNODHSVCYSDCFYHEKENQI 682
Db 959 SGOQHCELCIP-AHRDIDECILFGAEICKEGKC---VNTQPGYECYCKQGFY----- 1006
Qy 683 LHYDFSNLSSVGLMNGSPSTKGTKYFHFENISLCGHEGKKMALCTNNITDFTVKEIVA 742
Db 1007 --YDGNLLECV-----DVBCLDESN---CRNGVCE----- 1032
Qy 743 GSDDYTNLGAFCVOSTIIPSESKGFRAALSSQSITLADTFIGYTVETTLKNINIKEDMF 802
Db 1033 -----NTRGTYRCACT--PPAEYSPAQA-----QCLI-----PERW 1061
Qy 803 PVPTSQIPDVHFFYKSSATTATSCING-----RSTAVKMRCPN-TKSGAGVI 847
Db 1062 STPQRDYKCA---GASERTACVWGPWAGPALTFDDCCCRQPRLGTCRCPPRGTC--- 1115
Qy 848 SVPSKCPAGTCDGCTFYFLWESAEACPLC-----TEHDFHEIEGACKRGFQETLYVWNE 901
Db 1116 --SQCPSTQSESNF---WDTN---PLLLGKSPRDESDSESDSECR----- 1154
Qy 902 PKWCIKGISLPEKKLATCE 920
Db 1155 ---CVSGRCVPRPGAVCE 1170

RESULT 4

PCT-US95-02251-3
; Sequence 3, Application PC/TUS9502251
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR STIMULATING BONE
; TITLE OF INVENTION: CELLS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston

```
STATE: Texas
COUNTRY: United States of America
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
SOFTWARE: PatentIn Release #1.0, Version
SOFTWARE: #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/02251
FILING DATE: 18-FEB-1994
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/316,650
FILING DATE: 30-SEP-1994
CLASSIFICATION:
APPLICATION NUMBER: US 08/199,780
FILING DATE: 18-FEB-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: OMIC009P--
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679
TELEX: 79-0924
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1251 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-02251-3

Query Match 2.9%; Score 155.5; DB 5; Length 1251;
Best Local Similarity 20.2%; Pred. No. 5.6e-05;
Matches 149; Conservative 68; Mismatches 245; Indels 277; Gaps 48;

Qy 289 ECFCKP--GTFNKPQSFNCQVCP--NTYSEKAKCIRCKDDSQSGSSECTERPPC 344
Db 600 EAEPGCGKGCIMNTGSSYNCH--CNRGYRLHVGAGGRSCV-----DLNECAKPHLC 649
Qy 345 TTKDY---FQIHTPCDEBKGQIMY-----KWIEPKICREDLTDAIRLP---PSGEKKDC 393
Db 650 GDGFCINFPGHYKCN-----CYPGYRLKASRPIC-EDI-DECRDPSTCPDGKCNK 700
Qy 394 PCNPFGYNNSSSCHPCPPCTGSDGTEKCR---PCPAGTEPALGFYKWNVLPGNMKT 450
Db 701 P-----GSFKCIACQPGYRSGGGACRDVNECSGTPCSPG---WCEKLPGSYRC 747
Qy 451 SC-----FNVGNSKC---DGMNGHEVAGDHQSGAGGSDNDYLILNLHIPGFKPPTSMTG 502
Db 748 TCAOIRTRGLSCIDVDDCEAGVKODGICTNTPGFCOCLSGYHL-----GQTK 546
Qy 503 ATGSELGRITFV---FETLC--SADCVLYFMVDINRKSTNVVSW-----GQTK 546
Db 797 --SRDRSCEDIDECDFFAACIGGDCI-----NTNGSYRCLCPLGHLRLVGRK 842
Qy 547 -----EKQAYTHIFKNATFTFT---WAFORTNOGONRR-FTN- 581
Db 843 CKKDIDECSDPGLCLPHACENLQSYVCVDEGFTLTQDHGCEVEEPQHHKCEYLNF 902
Qy 582 -DMYKIYSITATNAVGDVYSSACRACALGSESGSC---VPC-----PPGHVIE 626
Db 903 DDTVFCDSVLATNVQ---QBC-CCSLGAGW-GDHCEIYPCVPYVSSAEFSLVDPGKRLH 957
Qy 627 KETNOCKECPDPTLYSIHQ---VYKEACI---PCGPGSKNNQDHSVCYDCDFEHEKENQI 682
Db 958 SGOQHCELCIP-AHRDIDECILFNAIECKGKC-----VNSQPGYECYCKQGFY----- 1005
Qy 683 LHYDFSNLSSVGSGLMNGPSFTSGTKYKYPHFNFNISLCGHEGKKMALCTNNTDFTVKEIVA 742

Db 1006 --YDGNLLECV-----DVECLDES-----CRNGVCENTWTR----- 1035
Qy 743 GSDDYTNLVGAFVQSTIIPSEKSGFRAALSSQSITLADTIGVTVETTLKNINIKEDMF 802
Db 1036 -----LPCACT--PPAEYSPAQA-----QCLI-----PBRW 1059
Qy 803 PVPTSQIPDVHFFYKSSPTATTSCING-----RSTAVKMCNP-TKSGACVI 847
Db 1060 STPQRDVKCA-----GASEBTACVWGPWAGPALTFDDCCCRQPRLGTQCRCPPRGTG-- 1113
Qy 848 SVPSKCPAGTCDGCTFFYELWESAACPLC-----TEHDFHEIEGACKRGQETLYVWNE 901
Db 1114 ---SOCPTSQSESNF---WDTG---PULLGKSPRDESDSEDSDECR----- 1152
Qy 902 PKWCIKGISLPEKKLATCE 920
Db 1153 ---CVSGPCVPRPGAVCE 1168

RESULT 5
US-08-199-780-3
; Sequence 3, Application US/08199780
; Patent No. 5763416
; GENERAL INFORMATION:
; APPLICANT: Bonadio, Jeffrey
; APPLICANT: Goldstein, Steven A.
; TITLE OF INVENTION: Gene Transfer Into Bone Cells
; TITLE OF INVENTION: And Tissues
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; FILING DATE: 18-FEB-1994
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: OMIC:002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 320-7200
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1252 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-199-780-3

Query Match 2.9%; Score 155.5; DB 1; Length 1252;
Best Local Similarity 20.2%; Pred. No. 5.6e-05;
Matches 149; Conservative 68; Mismatches 245; Indels 277; Gaps 48;

Qy 289 ECFCKP--GTFNKPQSFNCQVCP--NTYSEKAKCIRCKDDSQSGSSECTERPPC 344
Db 601 EAEPGCGKGCIMNTGSSYNCH--CNRGYRLHVGAGGRSCV-----DLNECAKPHLC 650
Qy 345 TTKDY---FQIHTPCDEBKGQIMY-----KWIEPKICREDLTDAIRLP---PSGEKKDC 393
Db 651 GDGFCINFPGHYKCN-----CYPGYRLKASRPIC-EDI-DECRDPSTCPDGKCNK 701
```

```

: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/199,780
: FILING DATE: 30-SEP-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Parker, David L.
: REGISTRATION NUMBER: 32,165
: REFERENCE/DOCKET NUMBER: UMIC:008
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (512) 418-3000
: TELEFAX: (713) 789-2679
: TELEX: 79-0924
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1252 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-316-650-3

Query Match 2.9%; Score 155.5; DB 2; Length 1252;
Best Local Similarity 20.2%; Pred. No. 5.6e-05;
Matches 149; Conservative 68; Mismatches 245; Indels 277; Gaps

QY 289 ECFPCRP--GTFSPKPGSPNQVCPR--NTYSEKAKECIRKDDSQSPSGSSETERPPC 344
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 601 EAEPCPGPGKICMNTGSSYNCH--CNRGYRLHVGAGGRSCV-----DLNECAKPHLC 650
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 345 TTRDY---FOIHTPCDEEGKTOIY-----KWIEPKICREDLTDAILRP---PSGEKKDC 393
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 651 GDGGFCINFGHYKCN-----CYPGYRLKASRPIC-EDI-DECRDPSTCPDGKCNK 701
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 394 PPCNPFGFYNNSSSCHPCPPGTFSDGTEKCR---PCPAGTEPALGFYKWNVLPGNNKT 450
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 702 P-----GSFKCLACPGYRSGGGACRDVNECSEGTPCSPG----WCEKLPGSYRC 748
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 451 SC-----FNWGNKSC---DMNGWEVAGDHIOGAGGSDNDYLILNLHPIGPKPPTSMTG 502
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 749 TCAQGITRTRGLRSLCDVDDCEAGKVCQDGICTNTPGSFQCCLSGYHL-----797
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 503 ATGSELGRITFV---FETLC-SADCVLYFMVDINRKSTNVVSW-----GGTK 546
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 798 --SRDRSRCEIDECDFPAACIGDCI-----NTNGSYRCLCPGLHRLVGGKR 843
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 547 -----EKOAVTHIIKFNATFTT---WAFORTNQGDNRR-FIN- 581
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 844 CKXIDECSDPGLCLPHACENLOGSYVCVDEGFTLTQDQHGCEEVQPHHKECINLP 903
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 582 -DMVKYISTATNAVGVASSCRACALGSEQSGSSC--VPC-----PPGHYIE 626
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 904 DDTVFCDSVLATNVTQ---QEC--CCSLGAGW-GDHCEIYPCVPVSSAEFSLVDPGKRLH 958
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 627 KETNOCKECPDYIYLSHQ--VYGRKACI--PCGPGSKNNQDHSVICYDCFFEYHEKNQI 682
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 959 SGQOHCCLCP-AHRDIDECILAFGEAICEKGKC---VNSQPGYECYCKQGFY-----1006
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 683 LHVDFSNLSSVGLMNGPSTSGTKYFHFNFISLCGHEGKKMALCTNNITDFTVKEIVA 742
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 1007 --YDGNLLECV-----DVDECLDSN---CRNGCVENTWR----1036
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 743 GSDDYTNLVGAFCVQSTIIPSESKGFRALSSQSIIADFTFIVGTVETTLKNINIKEDMF 802
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 1037 -----LPCACT-PPAEYSQAQ--QCLI-----PERW 1060
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 803 PVPTSQIPDVHFFYKSTATTSCING-----RSTAVKMCNP-TKSGAGVI 847
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 1061 STFPQDVKCA---GASEERTACVNGPWAGPALITFDDCCCRPRLGTQCRPCPRGTG--1114
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 848 SVPSKCPAGTCDCCTFYFLWESAECAPLC-----TEHDFHEIEGACKRGQETLYVWNE 901
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Db 1115 ---SQCTTSQESNSF---WDYTS---PLLGLKSPRDESDSESDSECR-----1153
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :
QY 902 PKWCIKGISLPEKKLATCE 920
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| :

```

RESULT 6
US-08-316-650-3
; Sequence 3, Application US/08316650

```

US-08-316-6300-3
: Sequence 3, Application US/08316650
: Patent No. 5942496
: GENERAL INFORMATION:
: APPLICANT: Bonadio, Jeffrey
: APPLICANT: Roessler, Blake J.
: APPLICANT: Goldstein, Steven A.
: APPLICANT: Lin, Wushan
: TITLE OF INVENTION: METHODS AND COMPOSITIONS
: TITLE OF INVENTION: FOR STIMULATING BONE CELLS
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Arnold, White & Durkee
: STREET: P.O. Box 4433
: CITY: Houston
: STATE: Texas
: COUNTRY: USA
: ZIP: 77210
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/316,650
: FILING DATE: 30-SEP-1994
: CLASSIFICATION: 514

```

```
Db 1154 ---CVSGPCVPRPGAVCE 1169
RESULT 7
US-08-718-388-7
; Sequence 7, Application US/08718388
; Patent No. 6271362
; GENERAL INFORMATION:
; APPLICANT: MORIKAWA, MINORU
; APPLICANT: HARADA, NAOKI
; TITLE OF INVENTION: GENE ENCODING IgG Fc REGION-BINDING
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08718.388
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY JR, GERALD M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 0230-111
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2594 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-718-388-7
Query Match 2.8%; Score 151; DB 4; Length 2594;
Best Local Similarity 17.5%; Pred. No. 0.00047;
Matches 202; Conservative 103; Mismatches 346; Indels 502; Gaps 62;
QY 18 AEAPRRGRSPHPANICWALAGCOAAWAGDLPSS-----SRPLPPCQEKDYHFEYT 71
Db 765 AAAPNCSGRPCVEGCVC---LPGFVAGGACVPASSCGCTFGQLQAPGQEV----- 814
QY 72 ECDSSGSRWRVAIPNSAVDCSLGDPV--RGKECTFCASGEYLEMKNQV-----119
Db 815 -----WADELORRCTNGATHQVTRDQK---SCPAGERCSVQNGLLCCYPDRGT 863
QY 120 CSKCGEGTSLSGIKFDWEDELPAFAGFNIAFMFTVVGPSDRPDGCGNNSWIPIRGNYI 179
Db 864 CQSGDHPHYVSPDGRFD-----FMGTCTYL--LVG-----SCQNAALPAFRVL 906
QY 180 ESN--RDCTVSLIYAVHLKSGY---VFEFYQ---YVDNNI--FEEFTIQDQCEMDTT 230
Db 907 VENEHRGQTSYTRAVRVEARGVKVAVRREYPGQVLDVLDVLPFOAADQGVQVFRQ 966
QY 231 TDKWKV---LTDNGEWSH--SVMLKSGTNIYWRITGILMGSKAVKPVL-----VKNT 280
Db 967 RDAVVRTDFGLTVTDNARVTAKEVPSSYAEALCGLGNFNGDPADDLALRGCGGAANAL 1026
QY 281 IEGVAYTSECPCPKPGTFSNKPFGSCNQVCP-----RNT 314
Db 1027 AFGNSWQEE---TRPGCGATEPGD-----CPKLSILVAQQLQSKNECGTILADPKGPFREC 1078
QY 315 YSE---KGA-KECI--RCKDDSQFSSESSECTERPPCTTKDYFIHTPCDEEGKTOIMYKW 368
Db 1079 HSKLDPOGAVRDCVYDRCLLPQ--SG-----PLCDALATYA--AACQAAAGT--VHPW 1126
QY 369 IEPKICREDLTDAIRLPPSGEKKDCP-----PCNPGFYNNNG 404
Db 1127 RSEELC-----PLSCPPSHIYEACSYGCGPLSCGDLPPVPGCGSGECHEGCVDEGFALSG 1180
QY 405 SSSCHP-----C-----PPG--TF-----SDGTEKCRPCPCPAGTPEPALGFEY 438
Db 1181 -ESCLPLASCGCVHOGTYHPPGQTFYPPGCDLSLCHQCGGLVSCSESSCGPHEA----- 1234
QY 439 KWNVLPGNMKTSCFNVNSKCDGMNGWEVAGDHTQSGAGSDNDYLILNLHIPGFKPPT 498
Db 1235 ----CQPSGSGSLGCVAVGSSTC-----QASGD-----PIH--T 1261
QY 499 SMTGATGSELGRIITVFETLCSADCVLFVFWVDI NRKSTNVVSWGTEKQAYTHLI--PK 557
Db 1262 TFDGRRDEFMGTCVYVLAQTCTCTRGLIRFAVLQE---NV---AWG--NGRVSVTRVITVQ 1314
QY 558 NATTFE---TW-----AFORTNOQD-----NRRFINDMKIY 587
Db 1315 VANFTLRLEQROMKVTVNGVDMKLPVVLANGQIRASQHSQDVVIETDFGLRVAYDLV--Y 1372
QY 588 SITATNAVDSGVASSCRACA-----LGSEQSG-----SSCV--- 617
Db 1373 YVRVTVPGNYIQQMCGLCGNYNGDPKDDFKPNSQAGNANEFGNSWEEVVPDSPCLPPT 1432
QY 618 PCPPGHIYETKETNOCKECPDPTYLSTHOVYKGEACIPCGPGSKNNOHDSVCY----- 669
Db 1433 PCPPG---SEDCIPSHKCPPE---LEKRYQKEF--CGLLSPTGPISSCHKLVDPQGP 1483
QY 670 -SDCFYIIEKENQILHYDFSNLSSVGLMNGPSFTSKGTYKHFHFNISICGHEGKMLALC 728
Db 1484 LKDCIF-----DLCLGGGNLSILC 1502
QY 729 TNNITDFTVKEIVAGSDDYTNLVGAPV--CQST---IIPSESKGFRALSSQSIILADTF 783
Db 1503 SN-----IHAYVSACQAAAGHVEFWRTET-----1527
QY 784 IGVTVETTLKNINIKEDMFPVPTSQIPDVHFFYKSTATTSCINGKSTAVKMKRNPTRSG 843
Db 1528 -----CPMECPNPNHSHYELCADTCSLGC-----1549
QY 844 AGVISVPKCPAGTCDG--CTFYFLWESAECPL-----875
Db 1550 -SALSAPQCQCAEGCQCDGSGFLYNGOACVPICQCGCYHNGVYVEPQTVLIDNCROQ 1608
QY 876 -----CTEHD-----FHEI-----EGACKRG 891
Db 1609 CTCHAGKGMVCOEHSCKPGQVQCPQSGGILSCVTKDPCHGVTCRPOETCKEQQGGVCLPN 1668
QY 892 FOETLYVWNEPKW 904
Db 1669 YEATCWLWGDPHY 1681
RESULT 8
US-08-718-388-9
; Sequence 9, Application US/08718388
; Patent No. 6271362
; GENERAL INFORMATION:
; APPLICANT: MORIKAWA, MINORU
; APPLICANT: HARADA, NAOKI
; TITLE OF INVENTION: GENE ENCODING IgG Fc REGION-BINDING
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/718,388
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: MURPHY JR, GERALD M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 0230-111
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 5405 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-718-388-9

Query Match 2.8%; Score 151; DB 4; Length 5405;

Best Local Similarity 17.5%; Pred. No. 0.0015;

Matches 202; Conservative 103; Mismatches 346; Indels 502; Gaps 62;

QY 18 AEAPRRSPWPWAWICWALAGCAWAGDLPSS-----SRPLPCQEKDYHFEY 71
DB 765 AAAPNSGRPCVEGCVC---LPGFVAGGACVPASSCGCTFGQLQAPGQEV----- 814
QY 72 ECDSSGRWRVAIPNSAVDCSLPDEV--RGKECTFSCASGEYLEMKNV----- 119
DB 815 -----WADLCQRCTCNATHQVTCRDKQ---SCPAGERCSVQNGLLGCYDPRGT 863
QY 120 CSKCGEGTYSLGSGIKFDEWDELPAFNSIATFTVVGPSDRPDGCGNNSWIPRGNYI 179
DB 864 CGSGDHPYVSPDGRFD-----FMGTCTYL--LVG-----SCGONAALPAFRVL 906
QY 180 ESN--RDCTVSLIYAVHLKSGY---VFEEYQ---YVDNNI--FEEFFIIONDCQEMDTT 230
DB 907 VNEHURGSOTVYTRAVRVEARGVAVRREYPGQVLDVLOYLPFAADGQVQVFRQG 966
QY 231 TDKWVK---LFDNGEWGSH--SVMLKSGTNILYRTTGTILMGSKAVKPVL-----VKNT 280
DB 967 RDVAVRTDFGLTVTDWNAVRTAKVPSSVAAELCGLGNFNGDPADDDIALRGCGQAANAL 1026
QY 281 IEGVAYTSECFCKPGTFSNKPGESENQVCP-----RNT 314
DB 1027 AFGNSWQEE---TRPGCGATEPGD---CPKLDLSLVAQQLQSKNECGILADPKGPFREC 1078
QY 315 YSE--KGA-KECI--RCKDDSQFSSSECTERPCTTKDYQIHTPCDEEKTQIMYKW 368
DB 1079 HSKLPDQGAIVDCVDRCLLPQ--SG-----PLCDALATYA--AACOAAGAT--VHPW 1126
QY 369 IEPKICREDLTDAILRPPSGEKKDCP-----PCNPGFYNG 404
DB 1127 REELC-----PLSCPPIHSHVEACSYGCLSCGDLVPYGGCGSECHCEVCDEGFALSG 1180
QY 405 SSSCHP-----C-----PPG-TF-----SDGTEKCRPCPAGTEPALGFEY 438
DB 1181 -ESCLPLASCGVHQGTVHPPGQTEYPGGCDLCHCEGGLVSCSESSCGPHEA----- 1234
QY 439 KWNVLPGNMKTSCFVNGNSKCDGMNGEYAGDHTQSGAGSDNDYILNLHIPGKPT 498
DB 1235 -----COPSGSGSLGCVAGSSTC-----QASGD-----PHY--T 1261
QY 499 SMTGATGSELGRITVFETLCSADCVLFMVDINRKNSTNVVESWGGTKEQAYTHLI--FK 557
DB 1262 TFDGRRFDFMGTCVIVLAQTCTRGLHRFAVLQE---NV--AWG--NGRVSVTRVITVQ 1314
QY 558 NATFTF-----TW-----AFORTNOGQD-----NRRFINDMVKIY 587

DB 1315 VANFTLRLEQRQWKVTNGVDMKLPVVLANGQIRASOHSQSDVVIETDFGLRVAYDLV--Y 1372
QY 588 SITATNAVGVASSCRACA-----LGSEDSG-----SSCV--- 617
DB 1373 YVYVTPVPGNYQOOCGLGNYNGDPKDDFKPNGSQAGNANEFNGSWEEVVPDSCPPT 1432
QY 618 PCPPGHYIEKETNOCKECPDPTVLSLTHOVYKKEACIPCGPGSKNNODHSVCY----- 669
DB 1433 PCPPG---SEDCIPSHKCPPE---LEKYOKEEF--CGLLSPTGLSSCHKLVDPQP 1483
QY 670 -SDCFYHEKENOILHYDFSLSVSGMLNGPSTFKGTKYFHFFNLSLCGHEGKKMALC 728
DB 1484 LKDCIF-----DCLCGGNNLSLCL 1502
QY 729 TNNITDFTVKEIVAGSDDYTNLVGAFV--CQST---IIPSESKGFPRAALSSQSIIADTF 783
DB 1503 SN-----IHAYVSACQAGGHVEPMTETF----- 1527
QY 784 IGVTVETTLKINIKEDMEFPVPTSQIPDVHFFYKSSSTATTSCTINGRSTAVKMRNPTKSG 843
DB 1528 -----CPMECPNPNHYELCADTCSLGC----- 1549
QY 844 AGVISVPSKCPAGTCDG--CTFYFLWESAECPL----- 875
DB 1550 -SALSAPPQOCQCAEGCOCDSGLYNGQACVPQOCQCVHNGVYVEPQTVLIDMCRQ 1608
QY 876 -----CTEHD-----FHEI-----EGACKRG 891
DB 1609 CTCHAGKGMVCEHSCCKPGQVCQPSGGILSCVTKDPCHGVTCTCRPQETCKEQGGVCLPN 1668
QY 892 FOETLYVWNEPKW 904
DB 1669 YEATCWLAGDPHY 1681

RESULT 9

US-08-365-470-3
Sequence 3, Application US/08365470
Patent No. 5632991
GENERAL INFORMATION:
APPLICANT: Glimbrone, Jr., Michael A.
TITLE OF INVENTION: Antibodies Specific For E-selectin And The Uses
TITLE OF INVENTION: Thereof
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX
STREET: 1100 New York Ave., NW
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/365,470
FILING DATE: herewith
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/102,510
FILING DATE: 05-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/850,802
FILING DATE: 13-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Markowicz, Karen R.
REGISTRATION NUMBER: 36,351
REFERENCE/DOCKET NUMBER: 0627.1350003
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600

```
; TELFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 610 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
US-08-365-470-3

Query Match          2.8%; Score 149; DB 1; Length 610;
Best Local Similarity 19.6%; Pred. No. 7e-05;
Matches 136; Conservative 79; Mismatches 214; Indels 264; Gaps 40;

QY 29 WSPA--WI-----CCWALAGCAA-----WAGDLFSSSRRLPPC-----OEKDYH 67
DB 65 YSPSYIIGIRKVNVMVWVGTKPLTEAKNWAPEGNRRQKD-EDCVEIYIKREKDVG 123
QY 68 FEYTECDSSGSRWVAIPNSA---VDCSGLPDPVRGKECTFSCASGEYLEMKNQVCSKC 123
DB 124 MNW---DERCSKKKALCYTAACNTWTS---HGCVEITINNYTCKC 165
QY 124 GEGTYSLSGKFKDEWDELPAQFSNIATFMDTVVGPDSRDPD---CNNSSWIPRGNYIE 180
DB 166 DRGF---SGLKCEQIVNCTA-----LESPEHGLVCSH---PLGNF-- 200
QY 181 SNRDDCTVSLIYAVHLKSGYVFEYQYVDNNIFEFQNDQCEMOTTTDKWVKLTDN 240
DB 201 SYNSSCSISC-----DRGYL-----PSSMET-----MQCMSS 227
QY 241 GWMGSHSVMLKSGTNILYWRITGILMSKAVKPVLVKNITIEGVAVTS-----ECFPCK 294
DB 228 GWSA-----PIPCNV-VECDVATNPANGFVECFQ-N 258
QY 295 PGTF-S-NKPGSFNCQ-----VCPRTYSEKGAKECIRCKDDSQ-FSGSSECT 339
DB 259 PGSEFWNTTCTFDCGEGFELMGAQSLQCTSSGNWDNKPCTKAVTCRAVRQPQNGSVRCS 318
QY 340 ERP-----PC--TTKDYFOIHTPCDEEGKTQIMYKWIIEP-KICREDLTDAILRPPSG 388
DB 319 HSPAGEFTFKSCNFTCEGFMLOGPAQVECTTQ--GOWTOQIPVCEAFQCTALSNERG 376
QY 389 EKKDCPPCPNPGYNNNGSSSCHPCPGTSDGTEKCRPCPAGTEPALGFYKMNVLPGNM 448
DB 377 -YMNCLPSASGFRYGSCEFSCEGFLVKGSKRLQCGPTG-----EW-----DNE 421
QY 449 KTSFCNVNGSKCDGNGWNEVAGDHQISGAGGSDNDYLLNLHLPKPKPTSWTGATGSEL 508
DB 422 KPTCEAV---RCDVH-----QPPKGLVRCVCAHSPI 448
QY 509 GRITVFETLCSADCVLYFMVDINRKSTNVVSWGTEKQAYTHILFKNATFTTWAQ 568
DB 449 GE--FTYKSSCAFSCEEGF-----ELYGSTQ-----LE 474
QY 569 RTNQGODNRRTIN-DWVKIYTSITATNAVDGVASS-----CR-ACALGSESGSCVPC 619
DB 475 CTSQGWTEEVPSQVVKCSSLAVPGKINMSSGEPVFTVCKFACPEGWTLNGSAARTC 534
QY 620 -PPGHYIEKTNQCKECPDPTVLSIHQVYKPA 651
DB 535 GATGHW-----SGLLPTCEAPTESNIPLVAGLSA 563

RESULT 10
US-09-209-668-19
; Sequence 19, Application US/09209668A
; Patent No. 6114517
; GENERAL INFORMATION:
; APPLICANT: Monia, Brett P.
; APPLICANT: Xu, Xiaoxing S.
; TITLE OF INVENTION: METHODS OF MODULATING TUMOR NECROSIS FACTOR
; FILE REFERENCE: ISPH-0336
; CURRENT APPLICATION NUMBER: US/09/209,668A
; CURRENT FILING DATE: 1998-12-10

; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 19
; LENGTH: 610
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-209-668-19

Query Match          2.8%; Score 149; DB 3; Length 610;
Best Local Similarity 19.6%; Pred. No. 7e-05;
Matches 136; Conservative 79; Mismatches 214; Indels 264; Gaps 40;

QY 29 WSPA--WI-----CCWALAGCAA-----WAGDLFSSSRRLPPC-----OEKDYH 67
DB 65 YSPSYIIGIRKVNVMVWVGTKPLTEAKNWAPEGNRRQKD-EDCVEIYIKREKDVG 123
QY 68 FEYTECDSSGSRWVAIPNSA---VDCSGLPDPVRGKECTFSCASGEYLEMKNQVCSKC 123
DB 124 MNW---DERCSKKKALCYTAACNTWTS---HGCVEITINNYTCKC 165
QY 124 GEGTYSLSGKFKDEWDELPAQFSNIATFMDTVVGPDSRDPD---CNNSSWIPRGNYIE 180
DB 166 DRGF---SGLKCEQIVNCTA-----LESPEHGLVCSH---PLGNF-- 200
QY 181 SNRDDCTVSLIYAVHLKSGYVFEYQYVDNNIFEFQNDQCEMOTTTDKWVKLTDN 240
DB 201 SYNSSCSISC-----DRGYL-----PSSMET-----MQCMSS 227
QY 241 GWMGSHSVMLKSGTNILYWRITGILMSKAVKPVLVKNITIEGVAVTS-----ECFPCK 294
DB 228 GWSA-----PIPCNV-VECDVATNPANGFVECFQ-N 258
QY 295 PGTF-S-NKPGSFNCQ-----VCPRTYSEKGAKECIRCKDDSQ-FSGSSECT 339
DB 259 PGSEFWNTTCTFDCGEGFELMGAQSLQCTSSGNWDNKPCTKAVTCRAVRQPQNGSVRCS 318
QY 340 ERP-----PC--TTKDYFOIHTPCDEEGKTQIMYKWIIEP-KICREDLTDAILRPPSG 388
DB 319 HSPAGEFTFKSCNFTCEGFMLOGPAQVECTTQ--GOWTOQIPVCEAFQCTALSNERG 376
QY 389 EKKDCPPCPNPGYNNNGSSSCHPCPGTSDGTEKCRPCPAGTEPALGFYKMNVLPGNM 448
DB 377 -YMNCLPSASGFRYGSCEFSCEGFLVKGSKRLQCGPTG-----EW-----DNE 421
QY 449 KTSFCNVNGSKCDGNGWNEVAGDHQISGAGGSDNDYLLNLHLPKPKPTSWTGATGSEL 508
DB 422 KPTCEAV---RCDVH-----QPPKGLVRCVCAHSPI 448
QY 509 GRITVFETLCSADCVLYFMVDINRKSTNVVSWGTEKQAYTHILFKNATFTTWAQ 568
DB 449 GE--FTYKSSCAFSCEEGF-----ELYGSTQ-----LE 474
QY 569 RTNQGODNRRTIN-DWVKIYTSITATNAVDGVASS-----CR-ACALGSESGSCVPC 619
DB 475 CTSQGWTEEVPSQVVKCSSLAVPGKINMSSGEPVFTVCKFACPEGWTLNGSAARTC 534
QY 620 -PPGHYIEKTNQCKECPDPTVLSIHQVYKPA 651
DB 535 GATGHW-----SGLLPTCEAPTESNIPLVAGLSA 563

RESULT 11
US-09-009-490A-89
; Sequence 89, Application US/09009490A
; Patent No. 6300491
; GENERAL INFORMATION:
; APPLICANT: Bennett and Mirabelli
; TITLE OF INVENTION: Oligonucleotide Modulation
; TITLE OF INVENTION: of Cell Adhesion
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Office of Jane Massey Licata
; STREET: 66 East Main Street
```

```

1 RESULT 12
2 US-08-185-432-19
3 ; Sequence 19, Application US/08185432
4 ; Patent No. 5750652
5 ; GENERAL INFORMATION:
6 ; APPLICANT: Artavanis-Tsakonas, Spyridon
7 ; APPLICANT: Busseau, Isabelle
8 ; APPLICANT: Diederich, Robert J.
9 ; APPLICANT: Xu, Tian
10 ; APPLICANT: Matsuno, Kenji
11 ; TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND
12 ; TITLE OF INVENTION: ANTIBODIES, ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
13 ; NUMBER OF SEQUENCES: 23
14 ; CORRESPONDENCE ADDRESS:
15 ; ADDRESSEE: PENNIE & EDMONDS
16 ; STREET: 1155 Avenue of the Americas
17 ; CITY: New York
18 ; STATE: New York
19 ; COUNTRY: U.S.A.

```

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/185,432
 FILING DATE: 21-JAN-1994
 CLASSIFICATION: 530
 ATTORNEY/AGENT INFORMATION:
 NAME: Misrock, S. Leslie
 REGISTRATION NUMBER: 18,872
 REFERENCE/DOCKET NUMBER: 7326-006
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 790-9090
 TELEFAX: (212) 869-8864/9741
 TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 19:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2703 amino acids
 TYPE: amino acid
 TOPOLOGY: unknown
 MOLECULE TYPE: protein
 US-08-185-432-19

Query Match 2.8% Score 148: DB 1: Length 2703:

Db 29 CDTRNP-GDKKCT-----RDECDTYFKVCLKEYQSRVTTAG-----PCSEFGSKS 72
QY 151 TFMOTVVGPSDSRPDCGNSSWIPRGNIYESNRDDCTVSLIYAVHLKKSQYVF-FEYQYV 209
Db 73 T---PVIG-----GN-----TFNLKYSRNEKNRIVIPFTFAMP 103
QY 210 DN-NIFFEFIQNDQCEMDTTTDKWKVLTJDNCEWGSVMLKSGTNILYVRT---TGI 264
Db 104 RSVTLLEAMDYNDNSTNDRIIEK-----ASHSGMINPSRQ---WQTLKHNTGA 150
QY 265 LMGSKAVKPVLVKNITIEGVAYTSECFP-CKPCTFSNKPFSFNCQVCPRNTYSEKAKEC 323
Db 151 AHFEYQIR-----VTCAEHYHFGCNKFCRP-----RDFETHHTCDQN-----GNKTC 194
QY 324 IRCKDDSQFSGSSECTE---RPPCTTKDYFIHTPCDEEGKTOIMYKW----- 368
Db 195 L-----EGWTG-PECNKAICRQCSPK-----HGCTVFGECRCQYWGQYCDKICIPHP 243
QY 369 -----IEP-----KICREDLTDAILRPPSGEKKDCPPCNPFGYNNSSSCHP 410
Db 244 GCYHGTCTEPWOCLETNWGGQLCDKDLNVCYGTHTPCLNGGTCSTGPDKY---QC-S 297
QY 411 CPPGTFSDGTKECR-----PCPAG---TEPALGPEYKWNVLPGNMTSCFNVGNSKCD 461
Db 298 CPEGYSQNCIEAIEAHACLSDPCHNGGSCLETSTGFE-----CV 335
QY 462 GMMGWVAG---DHIQS-----GAGGSDNDYLILNLHIPGFK--PPTSMTGATGSELG 509
Db 336 CAPGW--AGPTCTDNDIDCSPNCGHGGTCQDL-----VDGFKCICPPQWTGKT----- 382
QY 510 RITFVFETLCSAD-----CVLYPMVDINKS--TNVYESMGGTKEKOAYTHIIFKNAT 560
Db 383 -----COLDANECEGKPCV-----NANSCRNLIGSY-----YCDCI----- 413
QY 561 FTFTWAFORTNOGDNRRFINDMVKIYSITATNAVGVASSCRACALGSFQSSSC--- 616
Db 414 --TGWS-----GHNCIDIIND-----CR-----GOCQNGGSCRDLY 442
QY 617 ----VPCPPGH---YIEKETNOCKEP-----PDYLSIHQVYGEACIPG-PGSKN 661
Db 443 NGYRCICSPYAGDHCEKDKINECASNPMMNGHCQD-----EINGFQCLCPAGFSGNLC 496
QY 662 NQDHSVYCY-----SDCF-----FYHEKENOILHYDFSNLSSVG----- 594
Db 497 QLDIDYCEPNPCONGAOCFNLAWDYFCNCPED-----YEGKNCSHLKDHCRTPPCEVIDSC 552
QY 695 SLMNGPSFTSKGTFKFFHFNISLCGHEGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAP 754
Db 553 TVAVASNSTPEGVRY---ISSNVCGPHGK---CKSQ-----AG-----GRF 587
QY 755 VCOSTIIPSESQGFRA-----ALSSQSIILADTFIGVTVTT 791
Db 588 TCRC-----NKGFTGYCHENINDCESNPKNGGTCIDGVNSYKICSDGWEGTYCET- 640
QY 792 LKNINIKEDMFPVPTSQIPDVHFFYKSSSTATTSINGRSTAVKM-----RCNPTKSGAGY 846
Db 641 --NIN-----DCSKNPCHNGGTCRDLVNDFFCECKNGWKGTG 676
QY 847 ISVPSKCPAGTC-----DGTFFYL-----WESAEACPL-----CTEHDFHE-----I 884
Db 677 HSRDSQCDEATCNGGTCYDEGDTFKCMCPAGWEGA-TCNIARNSSCLPNPCHNGGTCVV 735
QY 885 EG-----ACKRGFOETLYWN-----EPKWCIKIGISLPERKKLATCETVDFWLKVAGVG 933
Db 736 SGDSFTCVCKEGWEGTCTQNTNDCSPHPCYNS-----GTCVDGDNWYRCECAPG 785

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 13:21:38 ; Search time 176.631 Seconds
(without alignments)
501.729 Million cell updates/sec

Title: US-10-073-333a-2

Perfect score: 5357

Sequence: 1 MLFRARGVRGWRGWRPAEA.....TCYFWKKNQKKKKTILNLFN 963

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 349150 seqs, 92025710 residues

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA.*

```
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB pep.*
```

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	5357	100.0	963	9	US-10-140-164-2
2	5307	99.1	1027	9	US-10-140-164-4
3	2796	52.2	1013	9	US-10-028-072-38
4	2796	52.2	1013	9	US-10-121-049-38
5	2796	52.2	1013	9	US-10-123-904-38
6	2796	52.2	1013	9	US-10-140-470-38
7	2796	52.2	1013	9	US-10-175-746-38
8	2796	52.2	1013	9	US-10-176-918-38
9	2796	52.2	1013	9	US-10-176-921-38
10	2796	52.2	1013	9	US-10-137-865-38
11	2796	52.2	1013	9	US-10-140-474-38
12	2796	52.2	1013	9	US-10-142-431-38
13	2796	52.2	1013	9	US-10-143-114-38
14	2796	52.2	1013	9	US-10-140-002-38
15	2796	52.2	1013	9	US-10-142-419-38
16	2796	52.2	1013	9	US-10-123-262-38
17	2796	52.2	1013	9	US-10-142-423-38
18	2796	52.2	1013	9	US-10-121-050-38
19	2796	52.2	1013	9	US-10-141-755-38

ALIGNMENTS

```
RESULT 1
US-10-140-164-2
; Sequence 2, Application US/10140164
; Publication No. US20030072736A1
; GENERAL INFORMATION:
; APPLICANT: Baker et al.
; TITLE OF INVENTION: Human Tumor Necrosis Factor Receptor Tr16
; FILE REFERENCE: PF514C1
; CURRENT APPLICATION NUMBER: US/10/140,164
; CURRENT FILING DATE: 2002-05-08
; PRIOR APPLICATION NUMBER: 09/637,856
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/148,348
; PRIOR FILING DATE: 1999-08-12
; PRIOR APPLICATION NUMBER: 60/148,683
; PRIOR FILING DATE: 1999-08-13
; PRIOR APPLICATION NUMBER: 60/148,870
; PRIOR FILING DATE: 1999-08-13
; PRIOR APPLICATION NUMBER: 60/148,758
; PRIOR FILING DATE: 1999-08-16
; PRIOR APPLICATION NUMBER: 60/149,181
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: 60/149,453
; PRIOR FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: 60/149,498
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 963
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-140-164-2
```

Query Match 100.0%; Score 5357; DB 9; Length 963;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 963; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLFRARGVRGWRGWRPAEAPRRGRSPWPSPAWICWALAGCAWAGDLPPSSSSRLPPP 60

DB 1 MLFRARGVRGWRGWRPAEAPRRGRSPWPSPAWICWALAGCAWAGDLPPSSSSRLPPP 60

```
QY 61 COEKDYHFEYTCDSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120
Db 61 COEKDYHFEYTCDSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120
QY 121 SKCGEGTSLGSGIKFDEWDELPAFNSIATFMDTVVGPDSRPGDGNSSWIPRGNYIE 180
Db 121 SKCGEGTSLGSGIKFDEWDELPAFNSIATFMDTVVGPDSRPGDGNSSWIPRGNYIE 180
QY 181 SNRDDCTVSLIYAVHLKSGYVFEYQYVDDNIFFEFQNDQCEMDTTTDKWKLTND 240
Db 181 SNRDDCTVSLIYAVHLKSGYVFEYQYVDDNIFFEFQNDQCEMDTTTDKWKLTND 240
QY 241 GEMGSHSVMLKSGTILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSCFPCKPGTFSN 300
Db 241 GEMGSHSVMLKSGTILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSCFPCKPGTFSN 300
QY 301 KPGSNCQVCPNTYSEKAKECIRCKDDSQFSGSSECTERPPCTTKDYFQIHTPCDEEG 360
Db 301 KPGSNCQVCPNTYSEKAKECIRCKDDSQFSGSSECTERPPCTTKDYFQIHTPCDEEG 360
QY 361 KTOIMYKWIETPKICREDLTDALRLPPSGEKKDCPCNPGFYNNSSCHPCPPGTFSDGT 420
Db 361 KTOIMYKWIETPKICREDLTDALRLPPSGEKKDCPCNPGFYNNSSCHPCPPGTFSDGT 420
QY 421 KECRCPAGCTEPALGFYKWNVLPNGMKTSFNVGNSKCDGMNGWEVAGDHIQSGAGS 480
Db 421 KECRCPAGCTEPALGFYKWNVLPNGMKTSFNVGNSKCDGMNGWEVAGDHIQSGAGS 480
QY 481 DNDYLILNLHIHIFGFPPTSMGTATGSELGRITFVETLCSADCVLYFMVDINRKSTNVVE 540
Db 481 DNDYLILNLHIHIFGFPPTSMGTATGSELGRITFVETLCSADCVLYFMVDINRKSTNVVE 540
QY 541 SWGGTKEKQAYTHIIFKNATFTFWAFQRTNQDNRREINDMKIYISITATNAVGVAS 600
Db 541 SWGGTKEKQAYTHIIFKNATFTFWAFQRTNQDNRREINDMKIYISITATNAVGVAS 600
QY 601 SCRACALGSEQSGSCVPCPPGHYIEKETNOCKECPDPTYLSIHQYVYKAEACIPCGPGSK 660
Db 601 SCRACALGSEQSGSCVPCPPGHYIEKETNOCKECPDPTYLSIHQYVYKAEACIPCGPGSK 660
QY 661 NNQDHSVCYSDCFFYHEKENQILHVDNLSVSGSLMNGPSTSKGTYKFFHFNISLGH 720
Db 661 NNQDHSVCYSDCFFYHEKENQILHVDNLSVSGSLMNGPSTSKGTYKFFHFNISLGH 720
QY 721 EGKKNALCTNNITDFTVKIIVAGSDDYTNLGVAFVCQSTIIIPSESKGFRALSSQSIILA 780
Db 721 EGKKNALCTNNITDFTVKIIVAGSDDYTNLGVAFVCQSTIIIPSESKGFRALSSQSIILA 780
QY 781 DTFIGYVTVETTLKINIKEDMFPVPTSQIPDVHFFYKSSATTSCINGRSTAVKMRNPT 840
Db 781 DTFIGYVTVETTLKINIKEDMFPVPTSQIPDVHFFYKSSATTSCINGRSTAVKMRNPT 840
QY 841 KSGAGVISVPSKCPACTCDGCTFYFLWESAECPLCTEHDHFIEGACKRGQFTLYVNN 900
Db 841 KSGAGVISVPSKCPACTCDGCTFYFLWESAECPLCTEHDHFIEGACKRGQFTLYVNN 900
QY 901 EPKWCIKGSLPEKKLATCETVDFWLKVGAGVGAFTAVLLVALTCYFVWKNKOKKKTIILN 960
Db 901 EPKWCIKGSLPEKKLATCETVDFWLKVGAGVGAFTAVLLVALTCYFVWKNKOKKKTIILN 960
QY 961 LFN 963
Db 961 LFN 963
```

RESULT 2

US-10-140-164-4

; Sequence 4, Application US/10140164

; Publication No. US20030072736A1

; GENERAL INFORMATION:

; APPLICANT: Baker et al.

; TITLE OF INVENTION: Human Tumor Necrosis Factor Receptor Tr16

; FILE REFERENCE: PF514C1

```
; CURRENT APPLICATION NUMBER: US/10/140,164
; CURRENT FILING DATE: 2002-05-08
; PRIOR APPLICATION NUMBER: 09/637,856
; FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/148,348
; PRIOR FILING DATE: 1999-08-12
; PRIOR APPLICATION NUMBER: 60/148,683
; PRIOR FILING DATE: 1999-08-13
; PRIOR APPLICATION NUMBER: 60/148,870
; PRIOR FILING DATE: 1999-08-13
; PRIOR APPLICATION NUMBER: 60/148,758
; PRIOR FILING DATE: 1999-08-16
; PRIOR APPLICATION NUMBER: 60/149,181
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: 60/149,453
; PRIOR FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: 60/149,498
; PRIOR FILING DATE: 1999-08-19
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-140-164-4
```

Query Match 99.1%; Score 5307; DB 9; Length 1027;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 953; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MLFRARGPVRGWGRPAEAPRRGRSPWPANICWALAGCAAAGWAGDLPSSSRPLPP 60
Db 1 MLFRARGPVRGWGRPAEAPRRGRSPWPANICWALAGCAAAGWAGDLPSSSRPLPP 60
QY 61 COEKDYHFEYTCDSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120
Db 61 COEKDYHFEYTCDSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120
QY 121 SKCGEGTSLGSGIKFDEWDELPAFNSIATFMDTVVGPDSRPGDGNSSWIPRGNYIE 180
Db 121 SKCGEGTSLGSGIKFDEWDELPAFNSIATFMDTVVGPDSRPGDGNSSWIPRGNYIE 180
QY 181 SNRDDCTVSLIYAVHLKSGYVFEYQYVDDNIFFEFQNDQCEMDTTTDKWKLTND 240
Db 181 SNRDDCTVSLIYAVHLKSGYVFEYQYVDDNIFFEFQNDQCEMDTTTDKWKLTND 240
QY 241 GEMGSHSVMLKSGTILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSCFPCKPGTFSN 300
Db 241 GEMGSHSVMLKSGTILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSCFPCKPGTFSN 300
QY 301 KPGSNCQVCPNTYSEKAKECIRCKDDSQFSGSSECTERPPCTTKDYFQIHTPCDEEG 360
Db 301 KPGSNCQVCPNTYSEKAKECIRCKDDSQFSGSSECTERPPCTTKDYFQIHTPCDEEG 360
QY 361 KTOIMYKWIETPKICREDLTDALRLPPSGEKKDCPCNPGFYNNSSCHPCPPGTFSDGT 420
Db 361 KTOIMYKWIETPKICREDLTDALRLPPSGEKKDCPCNPGFYNNSSCHPCPPGTFSDGT 420
QY 421 KECRCPAGCTEPALGFYKWNVLPNGMKTSFNVGNSKCDGMNGWEVAGDHIQSGAGS 480
Db 421 KECRCPAGCTEPALGFYKWNVLPNGMKTSFNVGNSKCDGMNGWEVAGDHIQSGAGS 480
QY 481 DNDYLILNLHIHIFGFPPTSMGTATGSELGRITFVETLCSADCVLYFMVDINRKSTNVVE 540
Db 481 DNDYLILNLHIHIFGFPPTSMGTATGSELGRITFVETLCSADCVLYFMVDINRKSTNVVE 540
QY 541 SWGGTKEKQAYTHIIFKNATFTFWAFQRTNQDNRREINDMKIYISITATNAVGVAS 600
Db 541 SWGGTKEKQAYTHIIFKNATFTFWAFQRTNQDNRREINDMKIYISITATNAVGVAS 600
QY 601 SCRACALGSEQSGSCVPCPPGHYIEKETNOCKECPDPTYLSIHQYVYKAEACIPCGPGSK 660
Db 601 SCRACALGSEQSGSCVPCPPGHYIEKETNOCKECPDPTYLSIHQYVYKAEACIPCGPGSK 660
```



```
RESULT 4
US-10-121-049-38
; Sequence 38, Application US/10121049
; Publication No. US2003002239A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C17
; CURRENT APPLICATION NUMBER: US/10/121,049
; CURRENT FILING DATE: 2002-04-12
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 38
; LENGTH: 1013
; TYPE: PRT
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 877, 882
; OTHER INFORMATION: unknown amino acid
US-10-121-049-38

Query Match      52.2%; Score 2796; DB 9; Length 1013;
Best Local Similarity 53.4%; Pred. No. 2.le-205;
Matches 494; Conservative 159; Mismatches 254; Indels 18; Gaps 10;

Qy 46 WAG---DLPSSSRRLPPCKQKDYHFEYTEDSSGSRWRVAIPNSAVDCSGLDPVGRKE 102
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 103 CTFCASGEYLEMKNQVCKSGEGTYSLGSGIKFDEWDELPAFNSNIATFMDTVVGPSSDS 162
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 91 CSFSCNAGEFLDMQDCKPCAEGRYSLGTGIRFDEWDELPHGFASLANMELDDSAES 150
Qy 163 RPDGNNSSWIPRGNVIESNRDCTVSLIYAVHLKKSQYVFEXQYVDNFIFFEFFIQND 222
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 151 -TGNCTSSKWPGRDGIYASNTDECTATIMYAVNLKQSGTVNFYYPDSIIFFEFFVND 209
Qy 223 OCQEMDTTDDKWKLTNDGNGSHSVMLKSGTNILYWRITGILMGSKAVKPVLYKNITIE 282
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 210 OCQP-NADDSRMWKTTEKG-WEFHSVELNRGNVLYWRITAFSVMTKVPKPVLYRNIAIT 267
Qy 283 GVAYTSECFCKPGTFSNKPSCFNCVCPRTYSEKGAKEICRCKDQSQFS--GSSECTE 340
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 268 GVAYTSECFCKPGTYADQKQSSCKLCUPANSYKNKGTSCHQC-DPDKYSEKSSSCNV 326
Qy 341 RPPCTTKDYFIHTPCDEEGTKIIMYKWIPEKICREDLTDAILRPPSGEKKDCPCPNPGF 400
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 327 RPACTDKDYFTYHTACDANGETQLMYKWKPKICSEDELEGAVKLPASGVKTHCPCPNPGF 386
Qy 401 YNNSSSCHPCPGTFFDGTKECRPCPAGTEPALGFYKWKNNVLPNGMKTSCFNVGSKC 460
Db      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 387 FKTNNSTCQPCPYGYSNGS-DCTRCRCPAGTEPAVGFEYKWKNNLTPTNNEITVLSGINFEY 445
```

US-10-140-470-38
; OTHER INFORMATION: unknown amino acid
; US-10-123-904-38
Query Match 52.2%; Score 2796; DB 9; Length 1013;
Best Local Similarity 53.4%; Pred. No. 2.1e-205;
Matches 494; Conservative 159; Mismatches 254; Indels 18; Gaps 10;
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3330R1C160
CURRENT APPLICATION NUMBER: US/10/140, 470
CURRENT FILING DATE: 2002-05-06
Prior Application removed - See Palm or File Wrapper
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 38
LENGTH: 1013
TYPE: prp
ORGANISM: Homo Sapien
FEATURE:
NAME/KEY: unsure
LOCATION: 877, 882
OTHER INFORMATION: unknown amino acid
US-10-140-470-38
Query Match 52.2%; Score 2796; DB 9; Length 1013;
Best Local Similarity 53.4%; Pred. No. 2.1e-205;
Matches 494; Conservative 159; Mismatches 254; Indels 18; Gaps 10;
Qy 46 WAG---DLPSSSRPLPPCKQEKYHFEYTCDSGSRWRVAIPNSAVDCSLGDPVVRGKE 102
Db 31 WAGTAFQVQTGTGPELHACKESYHYEYACDSTGSRWRVAVPHTPGLCTSLSDPVKCTE 90
Qy 103 CTFSASCSEYLEMKNQVCSKCGEGTSLGSGIKFDEWDELPAFNSIATFMDTVVGPDS 162
Db 91 CSFSCNAGEFLDMKDQCKPCAEGRYSLGTGIRFDEWDELPHGFASLSANMELDDSAES 150
Qy 163 RPDGNNSSWIPRGNYIESNRDDCTVSLIYAVHLKSKGYVFEYVDNNITFEFFIQND 222
Db 151 -TGNCSTSSKWPGRGDIASNTDECTATLMYAVNLKSGTGVNFYYPDSSIIFFEVQND 209
Qy 223 QCQEMDTTDDKWKVITDNGEWSHVMKSGTNILYWRITGILMGSKAVKPVLVKNITIE 282
Db 210 QCQP-NADDSRWMTTEKG-WEFHSELNRGNVLYWRITAFSVWTKVPKPVLRNIAIT 267
Qy 283 GVAYTSECFPCPKGTFSENKPGFQVCPNRYSEKGAKECIRCKDDSOFS--GSSECTE 340
Db 268 GVAYTSECFPCPKGTFYADKQGSFCKLCPANSYSNKGTSCHQC-DPDKYSKSGSSCNV 326
Qy 341 RPPCTTKDYFOHTPCDEEGKTQIMYKTEPKICREDLTDALRLPPSGEKKDCPCNPGF 400
Db 327 RPACTDKDYFYTHACDANGETQLMYKNAKPKICSEDLGAVKLPASGVKTHCPNPGF 386
Qy 401 YNNGSSCHPCPGTFSDGTEKCRPCAGTEPALGFYKWNVLPGNMKTSCFNVGNKSC 460
Db 387 FKTNNSTCQCPYGSYNGS-DCTRCPAGTEPAVGFEYKWNLTPTNMTTVLSGINFEY 445
Qy 461 DGMNGEVAAGDHIOGAGGSDNDYLILNLHIIPGFRPPTS-MTGATGSELGRITFEVETLC 519
Db 446 KGMTGWEVAGDHIYTAAGASDNDMILLTLVVPGRPQSVMAADTENKEVARITFVETLC 505

US-10-123-904-38
Query Match 52.2%; Score 2796; DB 9; Length 1013;
Best Local Similarity 53.4%; Pred. No. 2.1e-205;
Matches 494; Conservative 159; Mismatches 254; Indels 18; Gaps 10;
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3330R1C160
CURRENT APPLICATION NUMBER: US/10/140, 470
CURRENT FILING DATE: 2002-05-06
Prior Application removed - See Palm or File Wrapper
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 38
LENGTH: 1013
TYPE: prp
ORGANISM: Homo Sapien
FEATURE:
NAME/KEY: unsure
LOCATION: 877, 882
OTHER INFORMATION: unknown amino acid
US-10-140-470-38
Query Match 52.2%; Score 2796; DB 9; Length 1013;
Best Local Similarity 53.4%; Pred. No. 2.1e-205;
Matches 494; Conservative 159; Mismatches 254; Indels 18; Gaps 10;
Qy 46 WAG---DLPSSSRPLPPCKQEKYHFEYTCDSGSRWRVAIPNSAVDCSLGDPVVRGKE 102
Db 31 WAGTAFQVQTGTGPELHACKESYHYEYACDSTGSRWRVAVPHTPGLCTSLSDPVKCTE 90
Qy 103 CTFSASCSEYLEMKNQVCSKCGEGTSLGSGIKFDEWDELPAFNSIATFMDTVVGPDS 162
Db 91 CSFSCNAGEFLDMKDQCKPCAEGRYSLGTGIRFDEWDELPHGFASLSANMELDDSAES 150
Qy 163 RPDGNNSSWIPRGNYIESNRDDCTVSLIYAVHLKSKGYVFEYVDNNITFEFFIQND 222
Db 151 -TGNCSTSSKWPGRGDIASNTDECTATLMYAVNLKSGTGVNFYYPDSSIIFFEVQND 209
Qy 223 QCQEMDTTDDKWKVITDNGEWSHVMKSGTNILYWRITGILMGSKAVKPVLVKNITIE 282
Db 210 QCQP-NADDSRWMTTEKG-WEFHSELNRGNVLYWRITAFSVWTKVPKPVLRNIAIT 267
Qy 283 GVAYTSECFPCPKGTFSENKPGFQVCPNRYSEKGAKECIRCKDDSOFS--GSSECTE 340
Db 268 GVAYTSECFPCPKGTFYADKQGSFCKLCPANSYSNKGTSCHQC-DPDKYSKSGSSCNV 326
Qy 341 RPPCTTKDYFOHTPCDEEGKTQIMYKTEPKICREDLTDALRLPPSGEKKDCPCNPGF 400
Db 327 RPACTDKDYFYTHACDANGETQLMYKNAKPKICSEDLGAVKLPASGVKTHCPNPGF 386
Qy 401 YNNGSSCHPCPGTFSDGTEKCRPCAGTEPALGFYKWNVLPGNMKTSCFNVGNKSC 460
Db 387 FKTNNSTCQCPYGSYNGS-DCTRCPAGTEPAVGFEYKWNLTPTNMTTVLSGINFEY 445
Qy 461 DGMNGEVAAGDHIOGAGGSDNDYLILNLHIIPGFRPPTS-MTGATGSELGRITFEVETLC 519
Db 446 KGMTGWEVAGDHIYTAAGASDNDMILLTLVVPGRPQSVMAADTENKEVARITFVETLC 505
Qy 520 SADCVLYFMVDINRKSINNVSWGCTKQAYTHIIFKNATFTTFAFORTNOGDNRFF 579
Db 506 SYNCELYFMVGNSTNPTVETWKSQKQSYTIEENTTTSFTWAFORTFEHRSKY 565
Qy 580 INDVKKIYSITATNAVGVASSCRACALGSESGSCVPCPPGHVIEKTNCKECPDPT 639
Db 566 TNDVAKIYSINVTVMGVASVCRPCALEASDVGSCTSCPAGYIDRDSGTCCHSCPNT 625
Qy 640 YLSIHQVYKKEACIPCGPSKKNQDSVYSCDFYHEKENOILHYDFSNSVSGSLMNG 699
Db 626 ILKAHOPYGVQACVPCGPGTKNNKIHSLCYNDCTFSRNTPTFTFNFSALANTVTLAGG 685
Qy 700 PSFTSGKTYHFFNISLGHGKMKALCTNNITDTVKEIVAGSDDXTNLNGAFVQST 759
Db 686 PFETSKLYFHFFHLSLGCNGRKNVCTDNVTLRIPE--GESGFSKITAVVQAV 742
Qy 760 IIPSESKGFRALSSOSIILATFTGVVETTLKNINKEDMPPTSOIPDVHFFYKSS 819
Db 743 IIPPEVTGYKAGVSQVPSVADRLGILDTMDTLTGITSPALFLHLESIGIPDVIFFYRSN 802
Qy 820 TATTSCINGRSTAVKMRNPTKSGAGVISVPSKCPAGTCDCGTFEFLWESACPLCTEH 879
Db 803 DVTQSCSSGRSTIRVRCSPQKTVPGSLLPGLTCDGTCDCGNFHFLEWESAACPLCSVA 862
Qy 880 DPEIEGACKRGFOETLYVMNEPKWIKIGISLPEKKLATCETVDWLKYGAGVAGTAFTVL 939
Db 863 DYHAIYSSCVAGIQXTYYKREPKLCSGGISLPEBQRVTCTCKTIDFWLKVIGISAGTCTAIL 922
Qy 940 LVALTCYFWKKNOKKK----KTLN 960
Db 923 LTVLTCYFWKKNOKLEYKYSKLVN 947
RESULT 6

QY 520 SADCVLVFMVDINRKTNNVSWGGTKEKQAYTHIIIFKNATFTETWAFORTNOGODNRRF 579
Db 506 SYNCLELYFMVGNRSNTPTVETWKSQKQSYTYIEENTTSFTWAFORTFHEASRKY 565
QY 580 INDMVKIYSITATNAVGVASSCRALGSEGSQSCVPCPPGHYIEKETNOCKECPDPT 639
Db 566 TNDVAKIYSINTVMNGVASYCRPCALEASDVGSSTSCPAGYVIDRDSGTCHSCPNT 625
QY 640 YLSIHQVYKKEACIPCGPSKNNQDHSVCYDCFFYHEKENOILHYDFSNSLVSGSLMNG 699
Db 626 ILKAHQPYGVQACVPCGPGTKNNKIHSLCYNDCTFSRNTPTFTFNFYSALANTVTLAGG 685
QY 700 PSFTSGTKYFHHFNLISLGHGEGKMKALCTNNITDFTVKEIVAGSDDYTNLVGAFVQOST 759
Db 686 PSFTSKGLKYFHHFTLSLGNQGRKMSVCTDNVTLRIPE---GESGFSKSIATVYQAV 742
QY 760 IIPSESKGFRAALSSQSIILADFTFGTVVETTLKNINKEDMFPVPTSOIPDVHFFYKSS 819
Db 743 IIPPEVTGYKAGVSSQPVSLADRLIGVTTDMTLDGITSAPLEFLHLESIGIPDVIFFYRSN 802
QY 820 TATTSCINGRSTAVKMRNCPKTSKAGVISVPSKCPAGTCDGCTFVFLWESAACPLCTEH 879
Db 803 DVTQSCSSSRSTIRVRCSPQKTVPGSLLPCTSDGTCDCGNFHLWESAACPLCSVA 862
QY 880 DFHEIEGACKRGFOETLYVWNEPKWCIKISLPERKKLATCETVDFWLKVGAGVGAFTAVL 939
Db 863 DYHAIVSSCVAGIOXTVYXREPRLCSGSIPEQORVTICKTIDFWLKVGSAGTCTAIL 922
QY 940 LVALTCYFWKKNQKKK----KTILN 960
Db 923 LTVLTCYFWKKNQKLEYKYSKLVMN 947

RESULT 7

US-10-175-746-38
; Sequence 38, Application US/10175746
; Publication No. US20030027270A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Collin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C353
; CURRENT APPLICATION NUMBER: US/10/175,746
; CURRENT FILING DATE: 2002-06-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 38
; LENGTH: 1013
; TYPE: PRT
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 877, 882
; OTHER INFORMATION: unknown amino acid
US-10-175-746-38

Query Match 52.2%; Score 2796; DB 9; Length 1013;
Best Local Similarity 53.4%; Pred. No. 2.1e-205;
Matches 494; Conservative 159; Mismatches 254; Indels 18; Gaps 10;
QY 46 WAG---DLPSSSSRPLPPCOEKDYHFEYTECCSSSRWRVAIPNSAVDCGLDPVVRGE 102
Db 31 WAGTAFQVQTGGPELHACKESYHYEYACDSTGSRWRVAVPHTPGLCTSLSDPVKGT 90
QY 103 CTFSCASGEYLBKMQVCKSGEGTYSLSGSGIKFDEWDELPAFAGSNIATFMTVYVGPS 162
Db 91 CSFSGNAGEFLDMKQCKPCAEGRYSLGTGIRFDEWDELPHGFASLANMELSDSAES 150
QY 163 RPDGNNSSWIPRGNIIESNRDDCTVSLIYAVHLKSGVVFYQYVDNNIIEFEFIQND 222
Db 151 -TGNTCTSSKWWPRGDIIASNTDECTATLMYAVNLKQSGTVNEFYYPDSIIIFEVQND 209
QY 223 QCOEMDTTDDKWKLTDMGESHVMLKSGTNILYWRITTGILMGSKAVKPVLVKNITE 282
Db 210 QCOQ-NAADSRWMTTEKG-WEFHSELNRRGNVLYWRTTAFSVWTKVPKPVLRNIAT 267
QY 283 GVAYTSECPKPGTFSNKPQSFNCQVCPRNTYSEKGAKEICRCKDDSOFS--GSSECTE 340
Db 268 GVAYTSECPKPGTAYADKQSSFCFLCPANSYKNGETSCQC-DPOKYSEKSSCNV 326
QY 341 RPPCTTKDYEQIHTPCDEEGKTQIMVKWIEPKICREDLTDARLPSPGKCKDCPCNPGF 400
Db 327 RPACTDKDYFTHYACDANGETQOLMKWAKPKICSEDLEGAVKLPASGVKTHCPCNPGF 386
QY 401 YNNGSSCHPCPGTFSDBGTKECRPCAGTEPALGFYKWMNVLPNGMKTSCFNVGNSKC 460
Db 387 FKTNNSTQCPQPGSYSGS-DCTRCAGTEPAVGFEYKWNWLTPTNMTTVLSGINFEY 445
QY 461 DGMNGWEVAGDHIQSGAGSDNDYILNLHDPGKPPTS-WTGATGSELGRITTFVETLC 519
Db 446 KGMTGWEVAGDHIYTAAGASNDENMLTLVPGFRPPQSVMDTENKEVARITTFVETLC 505
QY 520 SADCVLVFMVDINRKTNNVSWGGTKEKQAYTHIIIFKNATFTETWAFORTNOGODNRRF 579
Db 506 SYNCLELYFMVGNRSNTPTVETWKSQKQSYTYIEENTTSFTWAFORTFHEASRKY 565
QY 580 INDMVKIYSITATNAVGVASSCRALGSEGSQSCVPCPPGHYIEKETNOCKECPDPT 639
Db 566 TNDVAKIYSINTVMNGVASYCRPCALEASDVGSSTSCPAGYVIDRDSGTCHSCPNT 625
QY 640 YLSIHQVYKKEACIPCGPSKNNQDHSVCYDCFFYHEKENOILHYDFSNSLVSGSLMNG 699
Db 626 ILKAHQPYGVQACVPCGPGTKNNKIHSLCYNDCTFSRNTPTFTFNFYSALANTVTLAGG 685
QY 700 PSFTSGTKYFHHFNLISLGHGEGKMKALCTNNITDFTVKEIVAGSDDYTNLVGAFVQOST 759
Db 686 PSFTSKGLKYFHHFTLSLGNQGRKMSVCTDNVTLRIPE---GESGFSKSIATVYQAV 742
QY 760 IIPSESKGFRAALSSQSIILADFTFGTVVETTLKNINKEDMFPVPTSOIPDVHFFYKSS 819
Db 743 IIPPEVTGYKAGVSSQPVSLADRLIGVTTDMTLDGITSAPLEFLHLESIGIPDVIFFYRSN 802
QY 820 TATTSCINGRSTAVKMRNCPKTSKAGVISVPSKCPAGTCDGCTFVFLWESAACPLCTEH 879
Db 803 DVTQSCSSSRSTIRVRCSPQKTVPGSLLPCTSDGTCDCGNFHLWESAACPLCSVA 862
QY 880 DFHEIEGACKRGFOETLYVWNEPKWCIKISLPERKKLATCETVDFWLKVGAGVGAFTAVL 939
Db 863 DYHAIVSSCVAGIOXTVYXREPRLCSGSIPEQORVTICKTIDFWLKVGSAGTCTAIL 922
QY 940 LVALTCYFWKKNQKKK----KTILN 960
Db 923 LTVLTCYFWKKNQKLEYKYSKLVMN 947

RESULT 8

US-10-176-918-38
; Sequence 38, Application US/10176918
; Publication No. US20030027275A1


```
QY 46 WAG---DLPSSSRPLPPCOEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLDPVVRGKE 102
Db 31 WAGTAFQVTOGTGPELHACKESYEHYEYACDSTGSRWRVAVPHTPGLCTSLSDPVKRGTE 90
QY 103 CTFSASGEYLEMKNQVCKSGEGTYSLSGKIFKDEWDELPAFNSIATFMDTVVGPSPDS 162
Db 91 CSFSCNAGBFLDMKDQCKPCAEGRYSLSGTGIRFDEWDELPHGFSASLANMELDDSAAES 150
QY 163 RPDGNNSSWIPRGNTIESNRDCTVSLIYAVHLKSGYVFYEQYVNDNIFEFFIQND 222
Db 151 -TGNCITSSKWWPRGDYIASNTDCTATLMYAVNLKSGTVNFEYYPDSIIFFEYVQND 209
QY 223 QCQEMDTTDDKVKLTNDGEMGSHVLMKSGTNIYLWRTTGILMGSKAVKPVLVKNITTE 282
Db 210 QCOP-NADDSRWKMTTEKG-WEFHVSVELNARGNVLWRTAFSVWTKVPKPVLRNIAIT 267
QY 283 GVAYTSECPCKPGTFSNKPFGSFCNOCVPRNTYSEKAKECIRCKDDSOFS--GSSECTE 340
Db 268 GVAYTSECPCKPGTYADKQGSFCKLCPANSYSNKGTSCHQC-DPDKYSEKGSSSCNV 326
QY 341 RPPCTTKDYFOIHTPCDEEGKTQIMYKWKTEPKICREDLTDALRLPSGKCKDCPCPCNGF 400
Db 327 RPACTDKDYFHTACDANGETQIMYKWKTEPKICREDLTDALRLPSGKCKDCPCPCNGF 386
QY 401 YNNGSSSCHPCPPGTSDGTEKCRPCPAGTEPALGFEYKWNVLPCNMKTSCFNVGNKSC 460
Db 387 FKTNNSTCQPCPYGSYNGS-DCTRCPCAGTEPAVGFEYKWNVLPCNMKTSCFNVGNKSC 445
QY 461 DGMNGWEVAGDHIQSGAGSDNDYILNLHIFGFKPPTS-MTGATGSELGRITTFVETLC 519
Db 446 KGMTGEVAGDHIYTAAGASDNDFMILTUVPGFRPQSVMDTENKEVARITTFVETLC 505
QY 520 SADCVLYFMVDINRKTNNVSVESWGGTKEQAYTHIFKNATFTFWAFORTNOGQDNRRF 579
Db 506 SYNCLEYFMVGNVSRNTPTVETWKSQKQSYTYIEENTTTSTFWAFORTTFHEASRKY 565
QY 580 INDMVKIYSITATNAVDGVASSCRACALGSEQSSGSCVPCPPGHYTEKTNCKECPDPT 639
Db 566 TNDVAKIYSINVTNMGVASYCRPCALEASDVGSSTSCPAGYYIDRSDGTCHSCPNT 625
QY 640 YLSIHQVYKEACIPGCGSKKNODHSVCYSDCFFYHEKENQILHYDFNLSSVGLMNG 699
Db 626 ILKAHQPGVQACVPCGPGTKNKHLSLCYNODTFSRNTPTTFNYSFALANTVTLAG 685
QY 700 PSFTSGTKYFHFHNISLCGHEGKKMALCTNITDPTVAKEIVAGSDDYTNLYCAFCQST 759
Db 686 PSFTSGKLYFHFHTLSLGCNGRKMVSCTDNVTLRIPE---GESGFKSITAYVQAV 742
QY 760 IIPSESKGFRAALSSQSIILADTFGVVETTLKNINIKEDMFPVPTSQIPDVHFFYKSS 819
Db 743 IIPPEVTGYKAGVSSQPVSLADRLIGVTTDMTLTGITSPAELFLESGLIPDVIFFRSN 802
QY 820 TATTSCINGRSTAVKMRNPTKSGAGVISVPSKPCAGTCGCTFFFLWESAACPLCTEH 879
Db 803 DYTQSSGSRSTIRVRCSPQNTVPSLULPGTCDGCTDCGNFHLWESAACPLCSVA 862
QY 880 DFHEIEGACKRGFQETLYYVNPBPKWIKGISLPEKKLATCETVDWLKYGAGVGAFTAVL 939
Db 863 DYHAIYSSCVAGIQXTYYXREPKLCSGISLPEQRVTICKTIDFLWLVGVSAGTCTAIL 922
QY 940 LVALTICYFKKKKQKK---KYLN 960
Db 923 LTVLTCYEWKKQKLEYKYKSKLVN 947
```

RESULT 10

US-10-137-865-38

```
; Sequence 38, Application US/10137865
; Publication No. US20030032155A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
```

```
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330RIC154
; CURRENT APPLICATION NUMBER: US/10/137,865
; CURRENT FILING DATE: 2002-05-03
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 38
; LENGTH: 1013
; TYPE: PRT
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 877, 882
; OTHER INFORMATION: unknown amino acid
US-10-137-865-38
```

Query Match 52.2%; Score 2796; DB 9; Length 1013;

Best Local Similarity 53.4%; Pred. No. 2.1e-205;

Matches 494; Conservative 159; Mismatches 254; Indels 18; Gaps 10;

```
QY 46 WAG---DLPSSSRPLPPCOEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLDPVVRGKE 102
Db 31 WAGTAFQVTOGTGPELHACKESYEHYEYACDSTGSRWRVAVPHTPGLCTSLSDPVKRGTE 90
QY 103 CTFSASGEYLEMKNQVCKSGEGTYSLSGKIFKDEWDELPAFNSIATFMDTVVGPSPDS 162
Db 91 CSFSCNAGBFLDMKDQCKPCAEGRYSLSGTGIRFDEWDELPHGFSASLANMELDDSAAES 150
QY 163 RPDGNNSSWIPRGNTIESNRDCTVSLIYAVHLKSGYVFYEQYVNDNIFEFFIQND 222
Db 151 -TGNCITSSKWWPRGDYIASNTDCTATLMYAVNLKSGTVNFEYYPDSIIFFEYVQND 209
QY 223 QCQEMDTTDDKVKLTNDGEMGSHVLMKSGTNIYLWRTTGILMGSKAVKPVLVKNITTE 282
Db 210 QCOP-NADDSRWKMTTEKG-WEFHVSVELNARGNVLWRTAFSVWTKVPKPVLRNIAIT 267
QY 283 GVAYTSECPCKPGTFSNKPFGSFCNOCVPRNTYSEKAKECIRCKDDSOFS--GSSECTE 340
Db 268 GVAYTSECPCKPGTYADKQGSFCKLCPANSYSNKGTSCHQC-DPDKYSEKGSSSCNV 326
QY 341 RPPCTTKDYFOIHTPCDEEGKTQIMYKWKTEPKICREDLTDALRLPSGKCKDCPCPCNGF 400
Db 327 RPACTDKDYFHTACDANGETQIMYKWKTEPKICREDLTDALRLPSGKCKDCPCPCNGF 386
QY 401 YNNGSSSCHPCPPGTSDGTEKCRPCPAGTEPALGFEYKWNVLPCNMKTSCFNVGNKSC 460
Db 387 FKTNNSTCQPCPYGSYNGS-DCTRCPCAGTEPAVGFEYKWNVLPCNMKTSCFNVGNKSC 445
QY 461 DGMNGWEVAGDHIQSGAGSDNDYILNLHIFGFKPPTS-MTGATGSELGRITTFVETLC 519
Db 446 KGMTGEVAGDHIYTAAGASDNDFMILTUVPGFRPQSVMDTENKEVARITTFVETLC 505
QY 520 SADCVLYFMVDINRKTNNVSVESWGGTKEQAYTHIFKNATFTFWAFORTNOGQDNRRF 579
Db 506 SYNCLEYFMVGNVSRNTPTVETWKSQKQSYTYIEENTTTSTFWAFORTTFHEASRKY 565
QY 580 INDMVKIYSITATNAVDGVASSCRACALGSEQSSGSCVPCPPGHYTEKTNCKECPDPT 639
```



```
Db 91 CSFSCNAGEFLDMKQSCPKCAEGRYSLGTGRFDEWDELPHGFASLSANMELDDSAAES 150
Qy 163 RPDGNNSSWIPRGNYIESNRDDCTVSLIYAVHLKSKGYVFEYOYVNNIIFEFEEFQND 222
Db 151 -TGNCSTSSKWVPRGDYIASNTDECTATLMYAVNLKQSGTVNFYYPDSSITFEFVQND 209
Qy 223 OCQEMDTTDDKWKLTIDNGEMGSHVMLKSGTNILYRTTGLMGSKAVKPKVLVKNITIE 282
Db 210 OCQP -NADDSRWKMTTEKG -WEEHSHVELNRGNVLYWRTTAFSVWTKVPKPKVLVKNIAIT 267
Qy 283 GVAYTSECFPPKPGTFSNKPQSFNCQVCPRTNYSBKGAKECIRCKDDSQFS--GSSECTE 340
Db 268 GVAYTSECFPPKPGTYADKQGSFCKLCPANSYKNGTSCHC -DPDKYSEKGSNCV 326
Qy 341 RPPCTTKDYFOHTPCDDEGKTQIMYKWIPIKICREDLTDAIRLPPSGEKKDCPCNPGF 400
Db 327 RPACTDKDYFTHTACDANGETQLMYKWKPKICSEDLGAVKLPASGVKTHCPCNPGF 386
Qy 401 YNNGSSCHPCPPGTFSDGTKECRPCPAGTEPALGFYKKNVLPCKMKTSCFNVGNKSC 460
Db 387 FKTNNSTQCPQPGYSGYNGS -DCTRCPAGTEPAVGFEYKKNWTLPTNMETTIVLSGINFEY 445
Qy 461 DGMNGWEVAGDHQSGAGSDNDYILNLHIFPGFKPPTS -MTGATGSELGRITTFVFETLC 519
Db 446 KGMTGWEVAGDHIYTAAGASDNDFMILTLLVPGFRPQSVMAADTENKEVARITFEVETLC 505
Qy 520 SADCVLVFWVDINRKSNTNVVSWGGTKKQAYTHIIFKNATFTTFAFORTNOGQDNRRF 579
Db 506 SVNCELYFMVGVNSRNTNPVETWKSQKQSYIIEENTTTSTFWAQTTFHEASRKY 565
Qy 580 INDMVKIYSITATNAVGVASSCRACALGSESGSCVPCPPGPHYIEKETNOCKECPPT 639
Db 566 TNDVAKIYSINVTNVNGVASYCRCALEASDVGSCTSCPAGYVIDRDSGTCHSCPNT 625
Qy 640 YLSIHQVYKKEACIPCGPSKKNQDHSVCYDCPFYHEKENOILHYDFSLSVGSGLMNG 699
Db 626 ILKAHOPYGVQACVPCGPGTKNKKTHSLUCYNDCTPSRNTPTFTFNFSALANTVTLAGG 685
Qy 700 PSFTSKGTGYHFFHNLISLCHGEGKMKALCTNITDFTVKEIVAGSDDYTNLGAFCVQST 759
Db 686 PSFTSKGLAYHFFHNLISLCHGEGKMKVCTDNTDLRIPE--GESGFSKITAVVQAV 742
Qy 760 IIPSEKSGPRAALSQSITLADTFGTGVVETTLKKNINKEDMPVPTSOIPDVHFFYKSS 819
Db 743 IIPPEVTGYKAGVQSPVSLADRLIGVTTDMTLDGITSAPAEFLHLESGLIPDVIFEYRSN 802
Qy 820 TATTSCINGRSTAVKMRCPNTKSGAGVISVPSKCPAGTCGCTFFYFLWESACPLCTEH 879
Db 803 DVTQSCSSGRSTIRVRCSPQKTVPGSLLLPCTGSDGTCGCGNFHFLWESAAACPLCSVA 862
Qy 880 DFHEIEGACKRGFQETLYVMNEPKMKIKGISLPEKKLATCTETVDFWLKYGAGVGAFTAVL 939
Db 863 DYHAIYSSCVAGIQXTYVYKREPKLCSGGISLPEQRVITCKITDFWLKVGISAGTCTAIL 922
Qy 940 LVALTCTYFKKKQKKK----KTLN 960
Db 923 LTVLTCTYFKKKQKLEYKYSKLYMN 947
```

RESULT 14

US-10-140-002-38

```
; Sequence 38, Application US/10140002
; Publication NO. US20030037623A1
```

GENERAL INFORMATION:

```
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
```

```
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C59
; CURRENT APPLICATION NUMBER: US/10/140_002
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See Palm or File wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 38
; LENGTH: 1013
; TYPE: PRT
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 877, 882
; OTHER INFORMATION: unknown amino acid
; US-10-140-002-38
```

Query Match 52.2%; Score 2796; DB 9; Length 1013;

Best Local Similarity 53.4%; Pred. No. 2.1e-205;

Matches 494; Conservative 159; Mismatches 254; Indels 18; Gaps 10;

```
Qy 46 WAG----DLPSSSRRLPPCOEKDYHFEYTECDSSSRWRVAIPNSAVDCSGLPDPVRGKE 102
Db 31 WAGTAFQTQGTGPELHACKSEYHYETACDSTGSRWRVAVPHTPGGLTSLSDVVKGTG 90
Qy 103 CFCSAGSEYLEMKNOVCSCGEGTYSLGSGIKFDEWDELPAFESNIATFMFTVTVGPSDS 162
Db 91 CSFSCNAGEFLDMKQSCPKCAEGRYSLGTGRFDEWDELPHGFASLSANMELDDSAAES 150
Qy 163 RPDGNNSSWIPRGNYIESNRDDCTVSLIYAVHLKSKGYVFEYOYVNNIIFEFEEFQND 222
Db 151 -TGNCSTSSKWVPRGDYIASNTDECTATLMYAVNLKQSGTVNFYYPDSSITFEFVQND 209
Qy 223 OCQEMDTTDDKWKLTIDNGEMGSHVMLKSGTNILYRTTGLMGSKAVKPKVLVKNITIE 282
Db 210 OCQP -NADDSRWKMTTEKG -WEEHSHVELNRGNVLYWRTTAFSVWTKVPKPKVLVKNIAIT 267
Qy 283 GVAYTSECFPPKPGTFSNKPQSFNCQVCPRTNYSBKGAKECIRCKDDSQFS--GSSECTE 340
Db 268 GVAYTSECFPPKPGTYADKQGSFCKLCPANSYKNGTSCHC -DPDKYSEKGSNCV 326
Qy 341 RPPCTTKDYFOHTPCDDEGKTQIMYKWIPIKICREDLTDAIRLPPSGEKKDCPCNPGF 400
Db 327 RPACTDKDYFTHTACDANGETQLMYKWKPKICSEDLGAVKLPASGVKTHCPCNPGF 386
Qy 401 YNNGSSCHPCPPGTFSDGTKECRPCPAGTEPALGFYKKNVLPCKMKTSCFNVGNKSC 460
Db 387 FKTNNSTQCPQPGYSGYNGS -DCTRCPAGTEPAVGFEYKKNWTLPTNMETTIVLSGINFEY 445
Qy 461 DGMNGWEVAGDHQSGAGSDNDYILNLHIFPGFKPPTS -MTGATGSELGRITTFVFETLC 519
Db 446 KGMTGWEVAGDHIYTAAGASDNDFMILTLLVPGFRPQSVMAADTENKEVARITFEVETLC 505
Qy 520 SADCVLVFWVDINRKSNTNVVSWGGTKKQAYTHIIFKNATFTTFAFORTNOGQDNRRF 579
Db 506 SVNCELYFMVGVNSRNTNPVETWKSQKQSYIIEENTTTSTFWAQTTFHEASRKY 565
Qy 580 INDMVKIYSITATNAVGVASSCRACALGSESGSCVPCPPGPHYIEKETNOCKECPPT 639
Db 566 TNDVAKIYSINVTNVNGVASYCRCALEASDVGSCTSCPAGYVIDRDSGTCHSCPNT 625
Qy 640 YLSIHQVYKKEACIPCGPSKKNQDHSVCYDCPFYHEKENOILHYDFSLSVGSGLMNG 699
Db 626 ILKAHOPYGVQACVPCGPGTKNKKTHSLUCYNDCTPSRNTPTFTFNFSALANTVTLAGG 685
```

Qy	700	PSFTSKGTGYFHFNISLCSGHECKKXWALCTNNITDFTVKEIVAGSDDYTNLYCAFPVCQST	759
Db	686	PSFTSKGLKYPHFHTLSLGNQGRKMSVCTDNDVLRUPE---GESGFSKSTAYVVCQAV	742
Qy	760	IIPSESGVFRALSSQSIILADFTGIVVETTLUKNINIKEDMFPVPTSPIDVHFHYKSS	819
Db	743	IIPPEVTGYKAGVSSQPVSLADRLIGVTTDMTLTGITSPAEFLHLESGLGIPDVIFFYRNS	802
Qy	820	TATTSCINGRSTAVKMKRNPTKSGAGVISVPSKCPAGTCGCGTFFYFLWESAECPLCTEH	879
Db	803	DVTQCSGSRSTTIRVRCSPQKTPGSLLLPGTCSDGTCGCGNFHLEWSAANCPLCSVA	862
Qy	880	DFHEIEGACRKGQETLYVWNEPKWCIKISLPEKKLATCEIVDFWKLKVGAGVGAFTAVL	939
Db	863	DYHAIVSSCVAGIOXTYVAXREPKLSCGSIPLPEQRVTICKTIDFWLKVGISAGTCTAIL	922
Qy	940	LVALTCYFWKKKQKK---KTLN	960
Db	923	LTVLTCYFWKKKQKLEYKYSKLVAN	947

```

RESULT 15
US-10-142-419-38
; Sequence 38, Application US/10142419
; Publication No. US20030044945A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C244
; CURRENT APPLICATION NUMBER: US/10/142,419
; CURRENT FILING DATE: 2002-05-10
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 38
; LENGTH: 1013
; TYPE: PRT
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 877, 882
; OTHER INFORMATION: unknown amino acid
US-10-142-419-38

```

Query Match 52.2%; Score 2796; DB 9; Length 1013;
Best Local Similarity 53.4%; Pred. NO. 2.1e-205;
Matches 494; Conservative 159; Mismatches 254; Indels 18; Gaps 10;

Qy	46	WAG----	DL	PSSSR	RP	LP	COEK	DY	H	F	E	T	C	O	S	S	G	R	W	V	A	P	N	S	A	V	D	C	S	G	L	P	D	P	V	R	G	K	E	102												
																																								103												
Db	31	WAG	A	P	O	V	T	O	G	T	E	L	H	A	C	K	E	S	E	H	E	T	A	C	D	S	T	S	R	W	V	A	P	H	T	P	G	L	T	S	L	D	S	D	V	K	G	T	E	90		
Qy	103	CT	S	C	A	G	E	Y	L	E	M	K	N	O	V	C	S	K	G	E	G	T	Y	L	S	G	S	K	T	F	D	E	W	L	P	A	G	F	N	I	A	T	F	M	T	V	V	G	P	D	S	162
Db	91	C	S	F	S	C	N	A	G	E	L	D	M	K	O	S	C	P	A	G	R	Y	S	L	G	T	R	F	D	E	W	L	P	H	G	F	A	S	L	S	A	N	M	E	L	D	S	A	A	E	150	

163	Qy	RPDCGNSSWIPRNGYTESNRDCTVSLIIYAVHLKXSGYVFFYQYVNNNIFPEFFIQND	222
151	Db	-TGNTCSKWKPGDGYIASNTDECTATLMTAVNLKQSGTYNFYFYPDSSIIPEFFVQND	209
223	Qy	QCQEDMTTDDKWKVLTDNGEWSHVMKSGTNILYWRITGILMGSKAVKPVLVKNITTE	282
210	Db	QCOP-NADDSERWKTTEKG-WEFHSVELNGRNVLWYRTTAFSVWTKVPKPVLVRIAIY	267
283	Qy	GVAYTSECFPCPKGTFSNKPSPFCNOCPVNTYSEKGAKECIRCKDDSQFS--GSSECTE	340
268	Db	GVAYTSECFPCPKGTGYADKQGSFCKLCPANSYNGKETSCHQC-DPDKYSEKSGSSCNV	326
341	Qy	RPPCTTKDYFOIHPPCDBEGKTQIWMYKWIIBPKICREDLTDAILRLPSGGEKKDCPPCNP	400
327	Db	RPACTDKDIYFTHACDANGETQIMYKWKAPKICSEDLGAVGLKPLASGVKTHCPCPNPGF	386
401	Qy	YNNSSSCHPCPPGTGTEKRCPCPAGTEPALGFYKWMNVLPGNMKTSCFNVGNSKC	460
387	Db	FKTWNSTCQPCPYGSYNGS-DCTRCPAGTEPAVGFEYKWNNTLPTNMETTVLSGINF	445
461	Qy	DGMNGEYVAGDHIOSGAGGSDNDYLILNLHLPQFKPPTS-MTGATSESLGRITVFETLC	519
446	Db	KGMTGWEVAGDHIYTAGASDNDFMIETLVVQFPRPQSVNADTENKEVARITFVETLC	505
520	Qy	SADCVLFWMDINRKSNTNVESNGGTEKEQAYTHIFPKNATFTFWAFORTNOQDNRRF	579
506	Db	SVNCELFVWGVNSRTNTPWTEKSGKGSYTVIEENTTTFTWAFORTTTFHEASRY	565
580	Qy	INDMKVIYSITATNAVDPGVASSCRACALGSEQSGSCVPCPPGHYTEKTNQCKECPDPT	639
566	Db	TNDVAKIYSINVTVMNGVASVCRPCALEASDVGSSTSCPAGYIDRDSGTCCHSCPWT	625
640	Qy	YLSHTQVYGRACITPCGPGSKNODHSVCYSDCFFYHEKENQILHYDFSMLWSVSGLMNG	699
626	Db	ILKAHQPYGVQACVPCPGPTGNKNIHSLCYNDCSTFRNTPTRTFNYSFALANTVTLAG	685
700	Qy	PSFTSKGTGFHFEENISLCGHEGKKNALCTNNTTDFTVKEIVAGSDDYTNLVCAFVCQST	759
686	Db	PSFTSKGLKFYEHFTLSLCNGQGRKNSVCHDNTDTRIPE---GEGFSKSIITAYVCQAV	742
760	Qy	IIPSESKFRAALSSQSIILADTFIGVTVETTLKNIKEDMPVPTSQIPDVHFYKSS	819
743	Db	IIPPEVNTGYKAGVSQVPSLADRLIGVTTDMLDGTSPAELEHLESGLIPDVIFVRSN	802
820	Qy	TATTSCINGRSTAVKMRNPTKSCAGVISVPSKCPAGCTDGCCTFYFLWESAECPLCTEH	879
803	Db	DVTQSCSGSRSTTIRVRCRSPKTVPGSLLLPCTGSDGTCDCGNFHFLLWESAAACPLCSVA	862
880	Qy	DFHIEGACRKRGFQETLYVMNEPKWCILKGISLPEKKLATCTVDVFWLKVYAGVGAFTAVL	939
863	Db	DYHALVSCVAGIOXTTVYVREPKLCSGGLSLPQRTVICTKIDFWLKVGISAGTCTAIL	922
940	Qy	LVALUTCYFWKKNQKK-----KTLIN 960	
923	Db	LTVALUTCYFWKKNQKLEYKYSKLVMN 947	

Search completed: May 12, 2003, 13:37:30
Job time : 179.631 secs

GenCore version 5.1.5
 Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 11:39:43 ; Search time 28.0673 Seconds
 (without alignments)
 3298.405 Million cell updates/sec

Title: US-10-073-333A-2

Perfect score: 5357

Sequence: 1 MLFRARGPVRGRGWRPAEA.....TCYFWKKKQKKKTLNLFN 963

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_73:*

1: pirl1.*

2: pirl2.*

3: pirl3.*

4: pirl4.*

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	242	4.5	1548	2	S34583
2	210	3.9	713	2	A35502
3	200	3.7	3635	2	T10053
4	198.5	3.7	677	2	C42125
5	194	3.6	5376	2	T42215
6	192.5	3.6	2823	2	F87908
7	192.5	3.6	2823	2	F87908
8	192.5	3.6	3102	2	T43291
9	183	3.4	1372	2	T25933
10	182	3.4	667	2	A48579
11	181	3.4	1766	2	A42125
12	178.5	3.3	3672	2	T23433
13	178.5	3.3	3704	2	T37316
14	177.5	3.3	1607	1	MMMSB2
15	177.5	3.3	2907	2	A57278
16	174.5	3.3	596	2	A45664
17	173.5	3.2	2871	2	A55624
18	173	3.2	1111	2	T26972
19	172	3.2	1786	1	MMMSB1
20	171.5	3.2	1620	2	T27283
21	170.5	3.2	2437	2	S42612
22	170	3.2	1557	2	T28811
23	168	3.1	1786	1	MMHUB1
24	166.5	3.1	557	2	A48434
25	166	3.1	1274	2	T42021
26	166	3.1	3002	2	A47217
27	165	3.1	1639	1	MMFFB2
28	164	3.1	2871	2	A55567
29	164	3.1	2918	2	A54105

30	164	3.1	3084	1	MMMSA
31	163	3.0	1297	2	T30274
32	161.5	3.0	1790	1	MMFFB1
33	161	3.0	3712	2	S18253
34	160.5	3.0	1609	1	MMHUB2
35	159.5	3.0	2352	2	T30201
36	156.5	2.9	3106	1	S53868
37	153.5	2.9	1680	2	A43434
38	152.5	2.8	1251	2	A57293
39	152	2.8	1820	2	A5494
40	152	2.8	3707	2	S18252
41	151.5	2.8	2824	2	T22759
42	150.5	2.8	2703	1	A24420
43	149	2.8	1203	2	A49175
44	149	2.8	1299	2	T43251
45	148.5	2.8	2491	1	A28372

ALIGNMENTS

RESULT 1

S34583

serine proteinase (EC 3.4.21.-) PC6B - mouse

C:Species: Mus musculus (house mouse)

C:Date: 02-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 05-Nov-1999

C:Accession: S34583

R:Nakagawa, T.; Murakami, K.; Nakayama, K.

FEBS Lett. 327, 165-171, 1993

A:Title: Identification of an isoform with an extremely large Cys-rich region of PC6,

A:Reference number: S34583; MUID:93327934; PMID:8335106

A:Accession: S34583

A:Status: Preliminary

A:Molecule type: mRNA

A:Residues: 1-1548 <NAK>

A:Cross-references: GB:D17583; NID:q407344; PIDN:BAA04507.1; PID:d1005033; PID:q44037

C:Keywords: hydrolase; serine proteinase

Query Match 4.5%; Score 242; DB 2; Length 1548;

Best Local Similarity 19.1%; Pred. No. 8.2e-08;

Matches 221; Conservative 97; Mismatches 382; Indels 456; Gaps 64;

QY	61	COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLDPDVRGKEC-----TFSCA 108
DB	347	CPGPHYHADKKRC-----RKAPN-CESCFG-----SHGDQCLSKCYFLNEETSSCV 394
QY	109	S-----GEYLEMKNQVCCKGEGTYSLGSGIKFDEWDELPAFGS-----NIATFMDTVVG 158
DB	395	TQCPDGSYEDIKKNVCGKSCENKAC---IGFHNCTECKGGLSLQSGRSCVTCEDGQFEN 451
QY	159	PSDSRP-----DGCNNSWIPRGNVIESNR--DCTVSLIYAVHLKSGGYVF 203
DB	452	GHDQCPCHRFATCSGAGADGGINCT---EGYVMEGRVCQSCVS--YYLDHSSEGGYK- 506
QY	204	FEYQVYDNNIIF-----EFFIQNDQOEMDTT 231
DB	507	SKCRDNSCLTCNPGFKNCSCPSGYLLDLGTOMGAICKDGEVIDDQGHQCTEASC 565
QY	232	DKWVKLTONGEWSHVMKSGTNILYNNRTTGLMGSKAVKPVLVKNITIEGVAYTSECF 291
DB	566	AKC-----WG-----PTQEDCISCPVTRVLDDGRC-----VMNCPSWKFPEKKQCH 606
QY	292	PKPGTFSNKPQSFNCVCPRTYSEKKAKEICRADD-----SQFSGSSECTERPPCTTK 347
DB	607	PC-----HYTCQGC-----QCSGFSNCTSCRADKHQGERFLYHGECLN---CPVG 649
QY	348	DY-FQIHT--PCDEGKQTOIMYKWIPEKTRCDLDAIRLPPS-----CE 389
DB	650	HYPAGHTCLPCPD--NCELCY---NPHICRCMGSYVIIPNHTCCKLECRQGEFQDSE 704
QY	390	KKDCPPCPNGFY---NNGSSSCHPCPGTFSQGTRECRPCAGTEPALGFYKKNVLPFG 446
DB	705	VEECMPCEGCLGCTEDDPGACTSCATGYIMFERHCYKACPEKT-----FGVKECRACG 759

laminin alpha-1 ch
 proteolisin - se
 laminin beta-1 cha
 laminin alpha-1 ch
 laminin gamma-1 ch
 Notch homolog prot
 laminin alpha-2 ch
 furin (EC 3.4.21.7
 latent transformin
 heparan sulfate pr
 hypohetical prote
 notch protein - fr
 Motch B protein -
 furin (EC 3.4.21.7
 insulin-like growt

Db	43	ECNCANTAPVNG--OCVDVNAEGPSKTLCPHSAKSAGCTQC CGNSFWMKDCQYSSGEGLP	101
Qy	136	----FDWEDEL-----PAGFSN-----IATPMDTVGPSPD---S	162
Db	102	SLCLSSDGDGVCTEAPGYPAPVCAAMTEQSVIACGDTTCVTIAAGNTYKGIADCAECS	161
Qy	163	RPDGCNNSWIPRGNYESNRDDCTVSLITYAVHLKSGYVFPEYQYVDNNIFPEFIQND	222
Db	162	APDA---TAGAEGKAVATCK--CGVS-----KYLKDNVCV-----D	193
Qy	223	OCQEMDTTDKWKVLTONGEWSHSUMLKSGTNILYWRITGILMGSKAVKPVLVKNITIE	282
Db	194	KAQCNSTGNKFVAV--DDSENGKNKVCSCSNLNL-----G	226
Qy	283	GVAVTSECFCKPCTFKNKPGSFNCQCPRNTYSEXGAK--ECIRCKDDDSQFSSE--C	338
Db	227	GVA-----NCDPFC-----SYDEQSKKIKCTKTDNNYLKTTSECTSC	263
Qy	339	TERPCCPTTKDYFOIHTPCDDEGKTQIMYKWIPEKICHEDLTDAILPPSGEKKKDCPCPNP	398
Db	264	VOKDQC--KDFG--FPKDDSS-----AGNK--CLPCND	290
Qy	399	GFYNGSSSCHPCP--PGTFSDGTKECRCPCPAGTEPALGFPEYKWNVVLPCNMKTKSCFNVG	456
Db	291	S--TDGIANCATCALVSGRSCAALVTCSDTGYKPSAD-----KTTCEAVS	335
Qy	457	NSKCDGMNGVEVACDHIHQSGAGGSDNDYLLIILHIIPGFKPPTSMTGATGSELGRITFVE	516
Db	336	NCKTPGCKACNECKENEVCTDCDGYTL-----TPTSQ-----	369
Qy	517	TLCSADCVLYPMVDINKNSTNVVSWGKTKEKOAYTHIIFKNATFTTWFQFT--NQGG	574
Db	370	--CIDSCA-----KIGNYATEGAKK-----LCKECTA---ANCKTCDQGG	407
Qy	575	DNRREFNDWKIYSITATNAVDGVASSCRACALGSEQSGSCVPCPPGHYIEK-----ET	629
Db	408	NOCKECPDPTYLSITHQVYKEACI--PCGPGSKNQDHSVCYSBCFFYHEKENQILHYDFS	688
Qy	438	SDCTECPGKALR-----YGGDTGKTGCEGCTGTGAGAC-----	473
Db	689	NLSVSGSLMNGPSTSGTKYFHFIFNLSLCHGHEKKMALCTNNITDFTVKEIVAG--SDD	746
Qy	474	--KTCGLTIDCASCYSECATTTEYPQNGVC---APKASRATPTCNDSPIQNGVCGTCADN	528
Db	747	YTNLVGAPVCOSTIIPESKGFRAALSQSIIILADTFIGVTVETTLKNINIKEDMPFVPT	806
Qy	529	YFKMNGG--CYETV---KYPGKTVCISAPN-----GGTCKAADGYKLDSTJTVCS	575
Db	807	SOIPDVHIFFYKSSATTATTCING--RSTAVKMRCN---PTKSGAGVISVPSKCPAGTCDCG	861
Qy	576	EGCKEC-----ASSDCTCTLDGYVKSASACTKCDASCTCNGA-----ATTCKAC	621
Db	862	--TFVFLWESAB--ACPLCTEHDHFHEIEGACKRGEQETLYVWNEPKWCIKGISLPEKK--LAT	918
Qy	622	ATGYKKTASGEGACTSC--ESDSNGVTGI-----KGCLNCAPPNPNKGSVL	665
Db	919	C-----ETVDFWLKVYAGVGAFTAVL-----LVALTCYFW	948
Qy	666	CYLIKDGSTNKSGLSTGALAGISVAVIVVGGIIGFLCWHE	707

RESULT 3
Tl0053
laminin alpha 5 chain - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 11-Jan-2000
C:Accession: Tl0053
R:Miner, J.H.; Lewis, R.M.; Sanes, J.R.
submitted to the EMBL Data Library, November 1997
A:Reference number: Z16923
A:Accession: Tl0053

Tl0053
 laminin alpha 5 chain - mouse (fragment)
 C:Species: Mus musculus (house mouse)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 11-Jan-2000
 C:Accession: Tl0053
 R:Miner, J.H.; Lewis, R.M.; Sanes, J.R.
 Submitted to the EMBL Data Library, November 1997
 A:Reference number: Z16923
 A:Accession: Tl0053

C:Species: Mus musculus (house mouse)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 11-Jan-2000
C:Accession: T10053
R:Miner, J.H.; Lewis, R.M.; Sanes, J.R.
submitted to the EMBL Data Library, November 1997
A:Reference number: Z16923
A:Accession: T10053

submitted to the EMBL Data Library, November 1997
A;Reference number: Z16923
A;Accession: T10053


```
QY 575 DNRFRINDMVKIYSITATNAVDCVASSCRACALGSEQSSGVCPGPHYIEKETNOCKE 634
Db 461 SSOA-----PIDGL-----CASEAQKAGNTCA-----NGVCTQ 488
QY 635 CPDPTYLSTHGVYKAEACIPGPGS---KNNODHSVCVSDCFYHEKENQILHYDFSNL 691
Db 489 CTNNFLYMGCGYSTQK-----APGFMCKTAGTGTGIC-----TEAANN----- 527
QY 692 SVGLMNGSFTSGTKYKPHFENISLCGHEGKKMALCTNITDFTVKEIVAGSDDTYN-L 750
Db 528 -----RYFVFGASNTDQSVLACSNPLGTLT-----GTGDTAKAY 562
QY 751 VGAFVQOSTIIPESKGFRAALSOSIILADTFIGVTVETTLKNIKEDMFPVPTSQIP 810
Db 563 VGVEGCSQCTAP-----AALSDGGM----- 582
QY 811 DVHFFYKSTATTSCINGRSTAVKRCNPTKSGAGVISVPSKCPAGTCGCTFYFLWESA 870
Db 583 -----APAVCTSCDSSK-----KPNRDGSGCV-----LCSVGCKSCV----- 615
QY 871 EACPLCTHDFEIEGACKRGFOETLYVWNEPKWICKIGISLPEKKLATCETVDFWLKVGA 930
Db 616 -----MDNIGECNSGF-----SLDNGKCVSSGANRSGLSAGA 648
QY 931 GYGAFATVAVL-----LVALTCYFW 948
Db 649 IAGISAVVAVVVGGLVAFLCWVF 671

RESULT 5
T42215
zonadhesin - mouse
N:Alternate names: sperm-specific membrane protein
C:Species: Mus musculus (house mouse)
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T42215
J:Gao, Z.; Garbers, D.L.
J.Biol. Chem. 273, 3415-3421, 1998
A:Title: Species diversity in the structure of zonadhesin, a sperm-specific membrane protein
A:Reference number: 222080; MUID:98123114; PMID:9452463
A:Accession: T42215
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-5376 <GAO>
A:Cross-references: EMBL:U97068; NID:G3327420; PID:G3327421; PIDN:AAC26680.1
C:Genetics:
A:Gene: Zan
A:Map position: 5
C:Function:
A:Description: functions in multiple cell adhesion processes
A>Note: found exclusively on the apical region of the sperm head
C:Keywords: cell adhesion

Query Match 3.6%; Score 194; DB 2; Length 5376;
Best Local Similarity 20.5%; Pred. No. 0.00053;
Matches 144; Conservative 63; Mismatches 204; Indels 290; Gaps 38;

QY 75 SSGRWRVAIPNSAVDCSLPDPVRGKCTECSACGEYLEMKNQVCSKCGEGYSLGSGI 134
Db 3256 SRGCTQSTCPAGATHCR-----NFKCPSTYCKNGDNGSSNCTEITLQCPNWS 3304
QY 135 KPEWDELPAFNSIATFMDTVVGPDSRPGDGNSSWIPRNYIESNRDDCTVSLIYAV 194
Db 3305 QFT--DCLP-----SCVPSCNRCEVTSVPSSCREG----- 3336
QY 195 HLKSGYVFEEYQYVDNNIFFEFTONDQCQEMDTTQKWKVLTDNGEWSHVMKSKST 254
Db 3337 -LCNHGFVFE-----DKCVPRTOCGCKDARGAIIPAG- 3368
QY 255 NILYWRITGILGSKAVKPVLYKNITIEGVAVTSCEFPCKPCTFSNKPQSFNQC--VCPR 312
Db 3369 -----KTWTSKGTQSCACV-----EGNIQCQNFQCP 3396
```

```
QY 313 NTYSEKAKECIRCDDSOFSGSSECTE---RPPCTTKDYFOIHTPC-----DEEG 360
Db 3397 ETY-----CKDNSE--GSSTCTKITLOCPAHTQ-----YTSLPSCLPSCLDPEG 3439
QY 361 KYQIMYKWIETPKI---CRE-----DLTDAI-RLPPSGEKKDCPP 395
Db 3440 -----LCKDISPRVPTCKEGVCQGYVLNSDKVLRABECDCKDAQGALIPAGKTWTSPG 3495
QY 396 CNPG-FYNNNGSSSCH--PCPPCTF---SDGTKECR---PCPAGTEPALGFEYKWNVL 444
Db 3496 CTQSCACMGAVQCOSSQCPGTCTYCKDNEDGNSCAKITLQCPAHS-----LFTNCL 3547
QY 445 PGNMKTSCFNVNGSKCDGMNGWEVAGDHIQSGAGSDNDYLILNLHIPGFKPPTSMTGAT 504
Db 3548 PPCL-PSCLDP-----DGL-----CKASPKVPSTCKE 3574
QY 505 GSELGRITFVETLGSADCVLYFMVDINKNSTN--VVESMGGTKEKQAYTHIFKNMATP 562
Db 3575 G-----CICQSGYVL-----SNKCLLRNRCGCKD--AHGALIPEDK--- 3609
QY 563 FTWAFQRTNQGDNRREFINDMVKIYSITATNAVDCVASSCRACALGSEQSSGVCPGPG 622
Db 3610 -TWVSR-----GCTQSC-VCTGSGIOCLSS--QCPPG 3637
QY 623 HYI---EKETNOCKECPDPTYLSIHQVYKGEACIPCGSGKNNQDH-----SVCYS 670
Db 3638 AYCKDNEDGSSNCARIPPOCPANSHYT---DCFPPCPSPSCDPEGHCEASGPRVLSTCRE 3694
QY 671 DCF---FYHEKENQILHYDFSLSVSGSLM-NGPSFTSKG 706
Db 3695 GCLCNPGFVLDKCVPRVECGCKDAQGALIPSGKTWTSPG 3735

RESULT 6
T23064
hypothenical protein T22A3.8 - Caenorhabditis elegans (fragment)
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000
C:Accession: T23064; T25096
R:Barlow, K.
submitted to the EMBL Data Library, October 1997
A:Reference number: 219669
A:Accession: T23064
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-2823 <WIL>
A:Cross-references: EMBL:AL008585; PIDN:CAA15432.1; GSPDB:GN00019; CESP:T22A3.8
A:Experimental source: clone H10E24
R:McMurray, A.
submitted to the EMBL Data Library, October 1996
A:Reference number: 219980
A:Accession: T25096
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-2823 <WIL>
A:Cross-references: EMBL:Z81125; PIDN:CAB03385.1; GSPDB:GN00019; CESP:T22A3.8
A:Experimental source: clone T22A3
C:Genetics:
A:Gene: CESP:T22A3.8
A:Map position: 1
A:Introns: 45/1; 282/2; 312/3; 416/2; 1255/3; 1418/3; 1776/2; 1988/2; 2760/2
C:Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like

Query Match 3.6%; Score 192.5; DB 2; Length 2823;
Best Local Similarity 18.8%; Pred. No. 0.00031;
Matches 205; Conservative 96; Mismatches 326; Indels 461; Gaps 59;

QY 39 LAGCOAAWAGDLPSSSRRLPPC-----QEKDY-----HFEYTECDSSGS 78
Db 819 LCACGQC---ECPSLDLNPPECISTELAVLGSVANSNEDNYVCINCPLEYGNKCEY--- 872
QY 79 RWRVAIPNSAVDCSG--LPDPVRGK--ECTF-----SCASGEYLEMKNQVCSKC- 123
```

```
Db 873 -----CSDGFEDPLTGKIECTCNGNIDPMGIGNCDS-----ETGKCLKCI 914
QY 124 GEGTYSLSGSIKFDWDELPAFNSIATFMDTVVGPSRDPGCCNNSSWIPRGNIENR 183
Db 915 GHTTGDSCESEKHHW-----GNAQLHT-----CKPGCHTQGA VNPQCSEENGE 959
QY 184 DCTVSLIYAVHLKKSgyvfeYQYVDNNIRFEFFIQNDQCOEMDTTDDKVKLTDNGEW 243
Db 960 CEKENYIGA-----IGSDCDQ-----VSGQC-NCKQGVFGKQCD 1019
QY 244 GSHSVNLKSGTNILYWRITGILMGSKAVKPVLVKNITIEGVAYTSECPCPKPGTFSN--- 300
Db 994 GS-----IGSDCDQ-----VSGQC-NCKQGVFGKQCD 1019
QY 301 --KPGSFN-----COVCPNTYSEKGAKECIRCKDDSQFSSESSECTERPPCT-----TKD 348
Db 1020 QCRPSYFNFTDAGCFCHCNIY---GSIEDGKC---DQTTGKCECENVEGTMCCKADG 1073
QY 349 YFOIHT-----PCDEEGKTQIMYKWIPEK-ICREDLTDAIRLPPSGEKKDCCPCNPGF 400
Db 1074 YFNITSGDCEDCGDPTGSEDVSCNLVTGQCVCKPGVT-----GLK--CDSCLPNF 1123
QY 401 Y---NNGSSSCHPCP-PGTFSDGTKECPCPAGTEPALGFEYKWNVLPGNMKTSFCFNVG 456
Db 1124 YGLTSEGTECEPCPAPGVQVCDPIDGSCVCPNPT-----VGMENCNTT-- 1167
QY 457 NSKCDGMNGWEVAGDHIOGAGGSDNDYLILNLHIPGKPPPTSMGTATGSELGRITFVEE 516
Db 1168 -----NAN-----DYHPLN-----GCK----- 1179
QY 517 TLCSADCVLYFMVDINRKNSTNVESWGKTKEQAY-----THIIFKNATFTFTWAFQ 568
Db 1180 -LC--DC-----SDIGSDGGMCTFTGCKCKAAVGLKCDLCTHGFFNFT----- 1223
QY 569 RTNQGODNRRFINDMVKIYSITATNAVGVASSCRACALGSQSGSSVCPGPPGHIYEKE 628
Db 1224 -----CEPC--GCNAAGTDPLOCKDGQCL--- 1245
QY 629 TNQCKECPDVTLSIHQVYKREACIPCGPG-----SKNQDHSVCYSDCF----- 673
Db 1246 CNEIGECF-----CKKNVHGK-CDQCGEGFTSLDSSNLKGCTECF--CFNRTSNCEQSD 1297
QY 674 -----FYHEKEN-OILHYDFSNLSS-----VGSMLNGPS 701
Db 1298 LVWQOQYAEADRAVPQEPWFEYTKKHINILLREKPSHSNYSPTDATPLYWPLPSTMLGDR 1357
QY 702 FTS-----KGTKYF-HFFNISLGHGKMKALCTNITDFTVK 738
Db 1358 TASYNGLRFLKIWNEDNRGLHGRDQOYFRHFPQVIFGNNRIELEHHPMEINDGDIY 1417
QY 739 EI-----VAGSDDYT-----NLVGAFCVOSTIIPSEKGFRAALSQSITIL 779
Db 1418 KIRLHESEWRVHRSPELTLTRKQMVVALQDTQGIYIRCTYTPARGDAINTQVSLDVAV 1477
QY 780 ADTFTGVTETLKNINIKEDM-----FPVPTSQIPDVHFYKSS----- 819
Db 1478 PESKIVAGLSTT-KAIGVEKCLGCPQGTGLSCQNPVEGYRKKHREYLNQADDIALIGW 1536
QY 820 TATTSCINGRTAVKMRNPTKSGAGVLSVPSKCPAGT-----CDGCTFYFLWESAE----- 871
Db 1537 SEPCSC-HGHS-----QTCNPD-----TSVCTDCBHNTFGDFCEHCLPGYIGDAREGGAN 1585
QY 872 -----ACPL-----CTEHDFHEIEGACKRGFOETLYVWNEPKWCIKG-ITSLPEKK 915
Db 1586 ACTKACACPLVENSFSFSCVAVDHGRGYVCNSCKPG-----YTGQYCEFCVAGYIGDPOHI 1640
QY 916 LATCETVD 923
Db 1641 GGTCSPCD 1648
```

RESULT 7

```
F87908
protein T22A3.8 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-Aug-2001
C:Accession: F87908; E87908
R:anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio
A:Reference number: A15000; MUID: 99069613; PMID: 9851916
A:Note: see websites genome.wustl.edu/gsc/c_elegans/ and www.sanger.ac.uk/Projects/C-
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A:Accession: F87908
A:Status: preliminary
A:Molecule type: DNA
A:Molecule type: DNA
A:Residues: 1-2823 <STO>
A:Cross-references: GB:chr-I; PIDN:CAA15432.1; PID:G3924779; GSPDB:GN00019; CESP:T22A
A:Accession: E87908
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2823 <ST2>
A:Cross-references: GB:chr-I; PIDN:CAB03385.1; PID:G3924881; GSPDB:GN00019; CESP:T22A
C:Genetics:
A:Gene: T22A3.8
A:Map position: 1
C:Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-lik

Query Match 3.6%; Score 192.5; DB 2; Length 2823;
Best Local Similarity 18.8%; Pred. No. 0.00031;
Matches 205; Conservative 96; Mismatches 326; Indels 461; Gaps 59;

QY 39 LAGCOAAWAGDLPSRSSRPLPPC-----OEKDY-----HFYEYTECDSSGS 78
Db 819 LGACQCC-----ECPSLDLNPNPECISTELAVLGSVASNEDNYVCINCPGLEGYKCKEY--- 872
QY 79 RWRVAIPNSAVDCSG--LPDPVVRGK--ECTF-----SCASGEYLEMKNOVCSSK- 123
Db 873 -----CSDGFEDPLTGKIECTCNGNIDPMGIGNCDS-----ETGKCLKCI 914
QY 124 GEGTYSLSGSIKFDWDELPAFNSIATFMDTVVGPSRDPGCCNNSSWIPRGNIENR 183
Db 915 GHTTGDSCESEKHHW-----GNAQLHT-----CKPGCHTQGA VNPQCSEENGE 959
QY 184 DCTVSLIYAVHLKKSgyvfeYQYVDNNIRFEFFIQNDQCOEMDTTDDKVKLTDNGEW 243
Db 960 CEKENYIGA-----IGSDCDQ-----VSGQC-NCKQGVFGKQCD 1019
QY 244 GSHSVNLKSGTNILYWRITGILMGSKAVKPVLVKNITIEGVAYTSECPCPKPGTFSN--- 300
Db 994 GS-----IGSDCDQ-----VSGQC-NCKQGVFGKQCD 1019
QY 301 --KPGSFN-----COVCPNTYSEKGAKECIRCKDDSQFSSESSECTERPPCT-----TKD 348
Db 1020 QCRPSYFNFTDAGCFCHCNIY---GSIEDGKC---DQTTGKCECENVEGTMCCKADG 1073
QY 349 YFOIHT-----PCDEEGKTQIMYKWIPEK-ICREDLTDAIRLPPSGEKKDCCPCNPGF 400
Db 1074 YFNITSGDCEDCGDPTGSEDVSCNLVTGQCVCKPGVT-----GLK--CDSCLPNF 1123
QY 401 Y---NNGSSSCHPCP-PGTFSDGTKECPCPAGTEPALGFEYKWNVLPGNMKTSFCFNVG 456
Db 1124 YGLTSEGTECEPCPAPGVQVCDPIDGSCVCPNPT-----VGMENCNTT-- 1167
QY 457 NSKCDGMNGWEVAGDHIOGAGGSDNDYLILNLHIPGKPPPTSMGTATGSELGRITFVEE 516
Db 1168 -----NAN-----DYHPLN-----GCK----- 1179
QY 517 TLCSADCVLYFMVDINRKNSTNVESWGKTKEQAY-----THIIFKNATFTFTWAFQ 568
Db 1180 -LC--DC-----SDIGSDGGMCTFTGCKCKAAVGLKCDLCTHGFFNFT----- 1223
QY 569 RTNQGODNRRFINDMVKIYSITATNAVGVASSCRACALGSQSGSSVCPGPPGHIYEKE 628
Db 1224 -----CEPC--GCNAAGTDPLOCKDGQCL--- 1245
```


A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1372 <MUR>
A:Cross-references: EMBL:U08015; PTDN:AAB37995.1; GSPDB:GN00022; CESP:W02C12.1
A:Experimental source: strain Bristol N2; clone W02C12
C:Genetics:
A:Gene: CESP:W02C12.1
A:Map position: 4
A:Introns: 29/1; 66/1; 774/2; 823/2; 1046/1; 1108/2; 1298/1

Query Match	3.4%	Score 183;	DB 2;	Length 1372;
Best Local Similarity	20.8%	Pred. No. 0.00056;		
Matches 154;	Conservative	75;	Mismatches 228;	Indels 284;
Gaps				
47;				
QY	65	DYHFEYTECDSSGR	---	WRVAIPNSAVDCSGLPDPVRGKECTFCAS-GEYLEMKNOV 119
DB	678	EYDVLVTATDNATNOACNFIRVKE	--	NCVDAADPVNGVQ---SCESWGP--OLKYKA 730
QY	120	CS-KCGEGTYSGLSGIKFDEDELPAGFSNIATFMDTVVGPDSRDPDCGNSSMIPRGNY 178		
DB	731	CSVECRDGF	-----	EFPRSPAVFYTCAA-----DG---KKPKNSP 763
QY	179	IESNR-DDCTVSLIYAVHLKSGVFEYQVDN	----	NIFFEFFIONDQCEMDTTDK 233
DB	764	STMFRYPOCT	-----	KHVPATKVIIVRIIYGSSPACTESSKEAFTKVQ--OTIDAIDSK 816
QY	234	W--VKLTD-NGEMGS	-----	HSVMLKS-GTNILY-----258
DB	817	WKMCSLTDANGCVGTQVRVCEGSGSLPEEGRRRRNPESVLASSFGVEIETPVKRRMLVD 876		
QY	259	---WRITT	-----	GIL-----MGSKAVK-----PVLVKNITIEG 283
DB	877	PSGLETTIRDALHNEILSGVLNFEKVLPNGRPDVGLSKIKEYLCQAGVVVRDL	----	932
QY	284	VAYTSECFCKPGTFSNKPSCFNOCVPRNTYSEKGA	K--	ECIRCKDDSQFSGS-----S 336
DB	933	---CVPCAGPY-HSATGECELCPIGEYQPLTARTCEFKCA-PQGITASEGAISEG 984		
QY	337	ECTER-PP	-----	CTTKDY-----FOIHTPCDECKTOIMYKWIPEKICRE 376
DB	985	ECKDNCPPGHQYDSLSDCVTCGVGYQPSAGAFEC-IPCG-IGKTTILSEFATSEDECDR 1042		
QY	377	DLTDAIRLPSEGEKKDCPPONPGYNN--GSSSCHPCPPGTFSDG	-----	419
DB	1043	ECPDGEOLASGV---CQPCQIGTYSRGENKKVCACPPGTTTATWSTREOCNTPCK 1099		
QY	420	---TKECRPCPAGTEPALGFEYK	-----	WWNVLPG-NKTKSCF-----453
DB	1100	PGQLVKETKNCQFCPRGT	-----	FQNEGEQSTCKLCPADHHTAAGPATAESQCFSTNQC 1154
QY	454	-----NVGSKDCGMNGWEVAGDHIQSGAGGSDNDYLIILNLHIPG 493		
DB	1155	ATGEYNCSWHANCIDLDPENDVPSYECRKPGRNGTHCTDAC	----	NDF-CLNDIGI--1207
QY	494	FKPPTSMWTGATGSELGRITFTFETLCSAD-CVLYFWVDINRK--STNVFWSWGTKKEQA 550		
DB	1208	-----CKKNIGNVEICKDHFSGDRCELRFQASNNKLWATIV	-----	GVV 1250
QY	551	YTHIIFKNATFTTWFORTNOQODNRREINDMWKIYSITATNAVDCVASSCRACALGSE 610		
DB	1251	VIGIIVIIIVFMSFRHHVQDNEKSSSTLADL	----	SPTANNILYTPPVCE-----1299
QY	611	QSGSSCVPCPP	---	GHYIEKE 628
DB	1300	-----PPRAFYYVYEDD 1311		

```

RESULT 10
A48579
trophozoite surface protein TSP11 - Giardia lamblia
C:Species: Giardia lamblia
C:date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 29-Jan-1999
C:Accession: A48579

```

R;Ey, P.L.; Khanna, K.K.; Manning, P.A.; Mayrhofer, G.
Mol. Biochem. Parasitol. 58, 247-257, 1993
A;Title: A gene encoding a 69-kilodalton major surface protein of Giardia
A;Reference number: A48579; MUID:93241215; PMID:8479449
A;Contents: Ad-1
A;Accession: A48579
A;Status: preliminary
A;Molecule type: nucleic acid
A;Residues: 1-667 <EY1>
A;Cross-references: GB:M95814; NID:gl59106; PID:gl59107
A;Note: sequence extracted from NCBI backbone (NCBIN:130056, NCBIP:130058)

Query Match	3.4%	Score 182:	DB 2:	Length 667:
Best Local Similarity	19.9%	Pred. No. 0.00028:		
Matches 146:	Conservative 58:	Mismatches 236:	Indels 292:	Gaps
QY	103	CTFCASGEYLEMKNVCSKCGEG-TYSLGSGIKFDBWDELPAQFSNIATFMDTVVGVPSD 161		
Db	141	CSAPKAGEENTPKAATCTCAAGFLHTPSEGUSSCE-ETCPEGY-----FGHTATAESK 194		
QY	162	SRPDGC--NNSSWIPRGNYIESNRDDCTVSLIYAVHLKKSGYVFQYVDNNIFFEFFIQ 220		
Db	195	KTKCSCTGGSEAPNVKGI-----GDCLKCM-----YNEASGNLT----- 229		
QY	221	NDQCEMDTTTDKWKLTLDNGENGSHISVMLKSGTNILYWRITG-----ILGSKA 270		
Db	230	--TCEKSAQKPKSLDKTSCNDG-----TGQNAFCSSSGDCEGCDGFIIDQCNC 279		
QY	271	VKPLVKRNITIEGVAYTSECFPCPGTFSNKPGEFNCVCPRNTYSEKGAKE-CIRCKDD 329		
Db	280	VK-----SDC-----KTENCKACT---NPKAANEVCTEIST 308		
QY	330	SQFSGSSECTERPPCTTKDYFQIHTPCDEEGKTQIMYKWIETPKICREDLTDAILRPPSGE 389		
Db	309	HLHTPTSQCVQY--COALGNVYAGTNADNK-----KACKE--CTVANCKTCD 352		
QY	390	KKDCPPCNPGEYNNGSS-----SCHPCPPGTESDGTKECRPCPA-----GTEPAL 434		
Db	353	QGOCOTCNDGYRNGDACSPCHESCKTCSAGTASD---CTECPGKALKYGNDGTKGTC 408		
QY	435	G-----FEYKWNVLPGN-----MKTSCFNVG 456		
Db	409	GEGCTTGGSGACKTCGLTIDGASYCEBDQNEYFONGICTSTTARTVATCKNS--NVA 466		
QY	457	NSKCDG-----MNG--WEVAGDHQSGAGSDNDYLILNLHIPGKPPKPTSMGTATGSE 507		
Db	467	NGICSSCTNGFLRMNGGCVETTRTKFPGKSVCEGANADADTKAPVPYGK-----VE 516		
QY	508	LGRITTFEFTLCSADCVLYFMVDINRKSTNVVESWGTKKEQAYTHLIFKNATFTFTWAF 567		
Db	517	AGKL-----VMCSKGC-----DTCSDATTC-----TKGDDGYTKI----- 546		
QY	568	QRTNQODNRRFIDMVKIYSHITATNAVGVASSCRACALGSSQSGSSCVPCPGHYH--- 624		
Db	547	-----EN-----SQCTKNC-----DSSCETCT---GAATTCKVCATGYKYKA 580		
QY	625	IBKETNOQCEPDPDTYLSIHQYVKRACIPCGPGSKNNODHSVCYSDCFYHEKENQIILH 684		
Db	581	LGEST--CTSCENDS-----NGVIGVRGCLNCAPPS----- 609		
QY	685	YDFSNLSSVSGSLMNGPSFTSKGPKFHFFHFNISLCGHEGKKMALCTNNITDFTVKEIVAGS 744		
Db	610	-----SSTGSQLV--CYLMMKGN-----TGGSVKNKSLGSTGAIGISVAAVVV--- 649		
QY	745	DDYTNLVGAFVC 756		
Db	650	--VAGLVG-FLC 658		

RESULT 11
A42125
trophozoite cysteine-rich surface antigen 170 - Giardia lamblia
N; Alternate names: CRP170; cysteine-rich surface antigen CRP170

Matches 116; Conservative 54; Mismatches 203; Indels 141; Gaps 30;

```

QY 279 ITIEGVAYTSECFPCPKP--GTFSENKPGSF--NCQVCPNRTYSEKGAKEICRCKDDSQFSG 334
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1413 LSCDCVAQGSSEFQCEYGGQCKCKFGVIGRRCERCAPGYN---PPECIKC-----QCNA 1465

QY 335 SSECTERP-PCITTKDYFIQHTPCDEEGKQTQIMYKWIETPKI-----CREDLTDAIRL--- 384
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1466 GQCDERTGQCFPPHVEQT-CDRCVSNAGFY---DPLIGCKQKCHPQSGEGGNLVCD 1521

QY 385 PPSGE-----KKDCPPCPCNPGFYNNSSSCHPCP---PGT---PSDGTK----- 421
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1522 PESGQCLRESMGGROCDCLAGFY--GPHCYGCSNCRAGTTTEICDATNAOCKCKENV 1579

QY 422 ---ECRPCPAGTEPALGFYKWNVLPGNMKTSFCFNVGNSKCDGMNGWEVAGDHIOGAG 478
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1580 YGGRCEACKAGT-----FDLSAENPL-GCVNCFCFGVTDSCRSSMYPVTIMSDMSFFLT 1633

QY 479 GSDN-----DYLILNLHIPG-----FKPPTSMTGATGSELGRITVFETL-----C 519
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1634 TDDNGMVDNKKDDTVIYTSEBTSNSVYFNVPYIEKKDYTTSYGLKLTFLKLTVPGRGKSM 1693

QY 520 SADCVLYFVWDINRKSTNV-VESWGG---TKEQOATHIIFKNATFTTFAFORTNOGOD 575
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1694 NAD-----ADVLGTGANMTIEYWEASEQPTNPEEQT-VKCKLVPENFLTAEGKTVTREE 1746

QY 576 NRRFINDMVKI-----YSITATNAVDGV--ASSCRACALGSQSGS 614
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1747 LMKVLHSLQNITLKASYFDHPKTKSTLYEGLSEIPNGVDSVIKASSVEQCCQCPAYTGP 1806

QY 615 SCVPCPPGHYIEKTNQCKECPDPTVLSLTHQVYKGE---ACIPC---GPGSKNNQDHQSV 668
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1807 SCOLCASGY-----HRVQSGFLGACVPCPCNGHSAATCDPDTGIC 1846

QY 669 YSDCFYHEKENOILHYDFSLSLSSVGLMNGPSF 702
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1847 -TDC-----EHTNGDHCFCNEGHYGNATNGSPY 1875

RESULT 13
T37316
C:Species: Caenorhabditis elegans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Jan-2000
C:Accession: T37316
R:Joh, K.; Zhu, K.; Hedgecock, E.M.; Inoue, T.; Hori, K.
A:Description: Laminin alpha chain gene in the nematode C. elegans.
A:Reference number: Z21681
A:Accession: T37316
A>Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-3704 <JOH>
A:Cross-references: EMBL:AB016806; PIDN:BAA32347.1
A:Experimental source: strain N2
C:Genetics:
A:Gene: epi-1
A:Map position: IV
A:Introns: 66/1; 284/3; 563/1; 1187/3; 1248/3; 1300/1; 1460/1; 1623/3; 2361/3; 2988/3; 3000/3
C:Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like F
Query Match 3.3%; Score 178.5; DB 2; Length 3704;
Best Local Similarity 22.6%; Pred. NO. 0.0036;
Matches 116; Conservative 54; Mismatches 203; Indels 141; Gaps 30;

QY 279 ITIEGVAYTSECFPCPKP--GTFSENKPGSF--NCQVCPNRTYSEKGAKEICRCKDDSQFSG 334
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1413 LSCDCVAQGSSEFQCEYGGQCKCKFGVIGRRCERCAPGYN---PPECIKC-----QCNA 1465

QY 335 SSECTERP-PCITTKDYFIQHTPCDEEGKQTQIMYKWIETPKI-----CREDLTDAIRL--- 384
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1466 GQCDERTGQCFPPHVEQT-CDRCVSNAGFY---DPLIGCKQKCHPQSGEGGNLVCD 1521

QY 385 PPSGE-----KKDCPPCPCNPGFYNNSSSCHPCP---PGT---PSDGTK----- 421
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1522 PESGQCLRESMGGROCDCLAGFY--GPHCYGCSNCRAGTTTEICDATNAOCKCKENV 1579

QY 422 ---ECRPCPAGTEPALGFYKWNVLPGNMKTSFCFNVGNSKCDGMNGWEVAGDHIOGAG 478
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1580 YGGRCEACKAGT-----FDLSAENPL-GCVNCFCFGVTDSCRSSMYPVTIMSDMSFFLT 1633

QY 479 GSDN-----DYLILNLHIPG-----FKPPTSMTGATGSELGRITVFETL-----C 519
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1634 TDDNGMVDNKKDDTVIYTSEBTSNSVYFNVPYIEKKDYTTSYGLKLTFLKLTVPGRGKSM 1693

QY 520 SADCVLYFVWDINRKSTNV-VESWGG---TKEQOATHIIFKNATFTTFAFORTNOGOD 575
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1694 NAD-----ADVLGTGANMTIEYWEASEQPTNPEEQT-VKCKLVPENFLTAEGKTVTREE 1746

QY 576 NRRFINDMVKI-----YSITATNAVDGV--ASSCRACALGSQSGS 614
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1747 LMKVLHSLQNITLKASYFDHPKTKSTLYEGLSEIPNGVDSVIKASSVEQCCQCPAYTGP 1806

QY 615 SCVPCPPGHYIEKTNQCKECPDPTVLSLTHQVYKGE---ACIPC---GPGSKNNQDHQSV 668
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1807 SCOLCASGY-----HRVQSGFLGACVPCPCNGHSAATCDPDTGIC 1846

QY 669 YSDCFYHEKENOILHYDFSLSLSSVGLMNGPSF 702
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1847 -TDC-----EHTNGDHCFCNEGHYGNATNGSPY 1875

RESULT 14
MMMSB2
N:Alternate names: laminin chain B2
C:Species: Mus musculus (house mouse)
C:Date: 28-Feb-1986 #sequence_revision 30-Jun-1991 #text_change 10-Dec-1999
C:Accession: A28469; A27729; A28082; S02680; S05327; S02037; A02870; S13544; S14552
R:Sasaki, M.; Yamada, Y.
J: Biol. Chem. 262, 17111-17117, 1987
A:Title: The laminin B2 chain has a multidomain structure homologous to the B1 chain.
A:Reference number: A28469; MUID:88059118; PMID:3680290
A:Accession: A28469
A:Molecule type: mRNA
A:Residues: 1-1607 <SAS>
A:Cross-references: EMBL:J03484; NID:g198694; PIDN:AAA39405.1; PID:g293688
R:Durkin, M.E.; Bartos, B.B.; Liu, S.H.; Phillips, S.L.; Chung, A.E.
Biochemistry 27, 5198-5204, 1988
A:Title: Primary structure of the mouse laminin B2 chain and comparison with laminin
A:Reference number: A27729; MUID:89000737; PMID:3167041
A:Accession: A27729
A:Molecule type: mRNA
A:Residues: 1-263; 'D', 265-336, 'C', 338-446, 'PS', 449-661, 'S', 663-885, 887-1155, 1157-1433
A:Cross-references: EMBL:J02930; NID:g198702; PIDN:AAA39408.1; PID:g293691
R:Ogawa, K.; Burbelo, P.D.; Sasaki, M.; Yamada, Y.
J. Biol. Chem. 263, 8384-8389, 1988
A:Title: The laminin B2 chain promoter contains unique repeat sequences and is active
A:Reference number: A28082; MUID:88228071; PMID:2836421
A:Accession: A28082
A:Molecule type: DNA
A:Residues: 1-215, 'A', 217-239 <OGA>
A:Cross-references: EMBL:J03749; NID:g198704; PIDN:AAA39409.1; PID:g554184
R:Fujiwara, S.; Shinkai, H.; Deutzmann, R.; Paulsson, M.; Timpl, R.
Biochem. J. 252, 453-461, 1988
A:Title: Structure and distribution of N-linked oligosaccharide chains on various dom
A:Reference number: S02678; MUID:88326259; PMID:2458101
A:Accession: S02680
A:Molecule type: protein
A:Residues: 227-238 <FUJ>
R:Hartl, L.; Oberhaeumer, I.; Deutzmann, R.
Eur. J. Biochem. 173, 629-635, 1988
A:Title: The N terminus of laminin A chain is homologous to the B chains.
A:Reference number: S00624; MUID:88225080; PMID:3267223
A:Accession: S05327

```



```
RESULT 15
A57278
fibrillin-2 precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 02-Aug-2002
C:Accession: A57278
R:Zhang, H.; Hu, W.; Ramirez, F.
J. Cell Biol. 129, 1165-1176, 1995
A:Title: Developmental expression of fibrillin genes suggests heterogeneity of extracellular matrix
A:Reference number: A57278; MUID:95263670; PMID:7744963
A:Accession: A57278
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2907 <ZHA>
A:Cross-references: GB:L39790; NID:G762830; PIDN:AAAT4908.1; PID:G762831
C:Superfamily: fibrillin 1; EGF homology
F:1239-1274/Domain: EGF homology <EGF1>
F:2488-2523/Domain: EGF homology <EGF>

Query Match 3.3%; Score 177.5; DB 2; Length 2907;
Best Local Similarity 20.1%; Pred. No. 0.0031;
Matches 224; Conservative 101; Mismatches 375; Indels 417; Gaps 67;

QY 41 GCOAAWAGDL-----PSSSRPLPPOCKDYHFEYTECDSSGSRW--VAIPNSAVDC 91
DB 17 GCVALLAOGTGDGQPPPPKTLWPQPPQVPAVAGSEGGFMGPYRDEGAVRASRVR 76
QY 92 SGLPDPVRGKE-C-----TFSCASGEYLEMKNO-----VC-SKCGEGTYS----- 129
DB 77 RGQETLRGPNVCGSRFHSCYCPGKWTLPGGNOCIVPICRNSCGDGFCSRPNMCTCSSQ 136
QY 130 -----LGSIGKFEDEWDELPAFNSIATFMDTVVGPDSRP-----DGCNN 169
DB 137 ISPTCGRKSIOQCSVRCMNGGTCAADHCCQCKY-----IGTCGQPVCEGNCQ 186
QY 170 SSWIPGNYIESNRDCTVSLIYAVHLKSGYVF-----FEYQVVDNIIFFEFIONDOCO 225
DB 187 G-----GRGICGNRCACV-----YGFPGQCEERDYRTGPGCTQ--VNNQMCQ 226
QY 226 EMDT-----TTDKVVKLTNDGEMSGHSMKSGTNILYWRITGILMGSAVKPVLVKNITIE 282
DB 227 GOLTGIVCTKTLCATIGRANGHPCMCFA-----OPQCRPGFIPNIRTG 272
QY 283 GVAYTSEC--FP--CKPGTFSNKPQSNQVCPNRTYSEKAKECTIRCKDDSQFSGSSBC 338
DB 273 ACQDVDECOAIPGLCOGGNCINTVGSFECR-CP-----AGHKQ-----SET 312
QY 339 TERPPCTTKDYFIHTPCDEEGKQIMYKWIPEKICREDLTDAIRLPPSGEKKDCPPCNP 398
DB 313 TK-----C-EDIDECVPIPGVCEGDC----- 334
QY 399 GFYNN--GSSSCHPCPPG--TFSDGTKECRPCAGTEPALGFYKWNVLPGNMKTS--- 451
DB 335 ---SNTVGSYFC-LCPRGVFTTDSR-CIDORAGTCFSLVNGRCAOELPGRMAKAQCC 389
QY 452 -----CFNVGN--SKC-----DGMNGWEVAGDHIQ--SGAGGSDNDYLIIN- 488
DB 390 CEPGRCSIGTIPEACPRGSEYRRLCLDGLPMGGIIPGSSVSRPCCGTGNSGNGYCPGT 449
QY 489 --LHIP--GFKPPTSMTCATSELGRI---TFVFET-----LC----- 519
DB 450 GFLPIPGDNGFSGVGGAGVGAGGOGPIITGLTILNIDICKKHANLCLNGRCIPTVSS 509
QY 520 -----SADCVLYFMVDINRKNSTNVESWGGTKKQAY---THIFKNATFTF 563
DB 510 YRCRCNMGYKQDANGDC-----IDVDECTSNPCNGDCVNTPGSYCKCHA----- 555
QY 564 TWAFORTNOCQ-----DNRRTINDMKIYSI-----TATNAV-DGV 598
DB 556 ---GFORTPTKQACIDIDEICQNGVLCKNGRCVNSDGSFQICNAGFELTDTGKNCVDHDE 613
QY 599 ASSCRACALG---SQSGSSCVPCPPGHYIEKETNQC---KEC-PPDIYLSIHQV--YKG 649
```

```
Db 614 CTTTNMCLNGMCINEDGSFKCV--CKPGFILAPNGRYCTDVEDCQTFGICMNGHCINNEGS 672
QY 650 EACIPCGPGSKNNQDHSVCY-----SDCFYHEKENQIILHYDFSNLSSVSGSLMNGPSF-- 702
Db 673 FRC-DCPPGLAVGVDGRVCDTHMRSTCYGEIKKGVCRVFPFGAVTKSECCCAN-PDYGF 730
QY 703 -----TSKGTKYFH-----FFNISLGHGEGKMMAL-----CTNNITDFTVKEIVAGSDD 746
Db 731 GEPQCPAPKNSAEFFHGLCSGGIGITVDGRDINECALDPPDICANGICE----- 778
QY 747 YTNLVGAFCQSTIIPSESKGFRAALSSQIILADTFIGVTVTTLKNINIKED----- 800
Db 779 --NLRTYRCNC-----NSGYEPDASGRNCIDIDECLE--VNRLLCDNGLCRNTPGSYS 827
QY 801 -----MFPVPTSOIPDVHFFYKSTATTSCINGRST-----AVKMRGNPTK--SGAGVI 847
Db 828 CTCPPGYVLPTEETCEDV-----NECESNPCVNGACRNNLGSFHCESPGSKLSSTGLI 882
QY 848 SVPSKCPAGTC-----DCCTFYF-----LWESAEACPLCTEHDHFHEIEG 886
Db 883 CIDSL--KGTWLNIOQNRCEVNINGATLKSECCATLGGAWGS--PCERC-----ELDA 932
QY 887 ACKRGFOETLYVWNEPKWCIKIGISLPEKKLATCETVD 923
Db 933 ACPRGFAR-----IKGV-----TCEDVN 950
```

Search completed: May 12, 2003, 13:21:11
Job time : 47.0673 secs



*

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 11:12:58 ; Search time 19,3568 Seconds
(without alignments)
2063.447 Million cell updates/sec

Title: US-10-073-333A-2

Perfect score: 5357

Sequence: 1 MLFRARGVVRGGRPAEA.....TCYFKKNQKKKTLNLN 963

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	242	4.5	1877	1	PCK5_MOUSE
2	240	4.5	1696	1	PCK5_BRACL
3	223.5	4.2	687	1	VS41_GIALA
4	210	3.9	713	1	TS44_GIALA
5	200	3.7	3718	1	LMA5_MOUSE
6	194	3.6	5376	1	ZAN_MOUSE
7	182	3.4	667	1	TS11_GIALA
8	178.5	3.3	3672	1	LML2_CAEEL
9	177.5	3.3	1607	1	LMG1_MOUSE
10	177.5	3.3	2307	1	FBM2_MOUSE
11	173.5	3.2	1581	1	LMG3_MOUSE
12	173.5	3.2	2871	1	FBM1_MOUSE
13	172	3.2	1786	1	LMB1_MOUSE
14	170.5	3.2	2437	1	NTC1_BRARE
15	170	3.2	1557	1	LM11_CAEEL
16	169	3.2	1639	1	LMG1_DROME
17	168	3.1	1786	1	LMB1_HUMAN
18	167.5	3.1	1587	1	LMG3_HUMAN
19	166	3.1	2871	1	FBM1_HUMAN
20	164.5	3.1	1790	1	LMB1_DROME
21	164	3.1	2871	1	FBM1_BOVIN
22	164	3.1	2871	1	FBM1_MOUSE
23	164	3.1	3084	1	LMAL_MOUSE
24	164	3.1	3110	1	LMAL_HUMAN
25	161	3.0	3712	1	LMAL_DROME
26	160.5	3.0	1609	1	LMG1_HUMAN
27	160.5	3.0	2911	1	FBM2_HUMAN
28	158	2.9	3695	1	LMA5_HUMAN
29	156.5	2.9	3106	1	LMA2_MOUSE
30	155.5	2.9	2003	1	NTC4_HUMAN
31	153.5	2.9	1680	1	FUR2_DROME
32	153.5	2.9	2491	1	MPRI_HUMAN
33	152	2.8	3707	1	PGBN_MOUSE

RESULT 1

PCK5_MOUSE

ID PCK5_MOUSE STANDARD; PRT: 1877 AA.

AC 004592; 062040;

DT 01-FEB-1995 (Rel. 31, Created)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Proprotein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)

DE (Proprotein convertase PC5) (Subtilisin/kexin-like protease PC5)

DE (Convertase PC5) (PC6) (Subtilisin-like proprotein convertase 6)

DE (SPC6).

GN PCSK5.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE OF 330-1877 FROM N.A. (ISOFORM PC5B).

RC STRAIN=ICR; TISSUE=Intestine;

RX MEDLINE=93327934; PubMed=8335106;

RA Nakagawa T., Murakami K., Nakayama K.;

RT "Identification of an isoform with an extremely large Cys-rich region

of PC6, a Kex2-like processing endoprotease.";

FEBS Lett. 327:165-171(1993).

RL [2]

RP SEQUENCE FROM N.A. (ISOFORM PC5A).

RC TISSUE=Brain, and Intestine;

RX MEDLINE=93324489; PubMed=8468318;

RA Nakagawa T., Hosaka M., Torii S., Watanabe T., Murakami K.;

RT "Identification and functional expression of a new member of the

mammalian Kex2-like processing endoprotease family: its striking

structural similarity to PACE4.";

J. Biochem. 113:132-135(1993).

RL [3]

RP SEQUENCE FROM N.A. (ISOFORM PC5A).

RC TISSUE=Adrenal cortex;

RX MEDLINE=93342056; PubMed=8341687;

RA Lussan J., Vieu D., Hamelin J., Day R., Chretien M., Seidah N.G.;

RT "cDNA structure of the mouse and rat subtilisin/kexin-like PC5: a

candidate proprotein convertase expressed in endocrine and

nonendocrine cells.";

Proc. Natl. Acad. Sci. U.S.A. 90:6691-6695(1993).

RL [4]

RP PARTIAL SEQUENCE, AND SUBCELLULAR LOCATION.

RX MEDLINE=97103178; PubMed=8947550;

RA De Bie I., Marcinkiewicz M., Malide D., Lazure C., Nakayama K.;

RT "The isoforms of proprotein convertase PC5 are sorted to different

subcellular compartments.";

J. Cell Biol. 135:1261-1275(1996).

RL [5]

RP DEVELOPMENTAL EXPRESSION.

RX MEDLINE=96293359; PubMed=8698813;

RA Constam D.B., Calfon M., Robertson E.J.;

RT "SPC4, SPC6, and the novel protease SPC7 are coexpressed with bone

P07207 drosophila
Q35516 mus musculus
Q28833 sus scrofa
Q01705 mus musculus
P17053 paramedius
Q9um47 homo sapien
P16581 homo sapien
Q16363 homo sapien
Q61982 mus musculus
P21783 xenopus lae
P08603 homo sapien
P25304 rattus norv

ALIGNMENTS

morphogenetic proteins at distinct sites during embryogenesis.";
[6]
J. Cell Biol. 134:181-191(1996).
DEVELOPMENTAL EXPRESSION.
MEDLINE=97436919; PubMed=9291583;
Rancourt S.L., Rancourt D.E.;
"Murine subtilisin-like proteinase spc6 is expressed during embryonic
implantation, somitogenesis, and skeletal formation.";
Dev. Genet. 21:75-81(1997).
-!- FUNCTION: LIKELY TO REPRESENT A WIDESPREAD ENDOPEPTASE ACTIVITY
WITHIN THE CONSTITUTIVE AND REGULATED SECRETORY PATHWAY. CAPABLE
OF CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF. MAY BE RESPONSIBLE
FOR THE MATURATION OF GASTROINTESTINAL PEPTIDES. MAY BE INVOLVED
IN THE CELLULAR PROLIFERATION OF ADRENAL CORTEX VIA THE ACTIVATION
OF GROWTH FACTORS.
-!- CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR
PROTEINS BY CLEAVAGE OF ARG-XAA-YAA-ARG-I-ZAA BONDS, WHERE XAA
CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.
-!- SUBCELLULAR LOCATION: PC5A IS SECRETED THROUGH THE REGULATED
SECRETORY PATHWAY. PC5B IS A TYPE I MEMBRANE PROTEIN LOCALIZED TO
A PARANUCLEAR POST-GOLGI NETWORK COMPARTMENT IN COMMUNICATION WITH
EARLY ENDOSOMES.
-!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS: PC5B/LONG (SHOWN HERE)
AND PC5A/SHORT; ARE PRODUCED BY ALTERNATIVE SPLICING.
-!- TISSUE SPECIFICITY: PC5A IS EXPRESSED IN MOST TISSUES BUT IS MOST
ABUNDANT IN THE INTESTINE AND ADRENALS. PC5B IS EXPRESSED IN THE
INTESTINE, ADRENALS AND LUNG BUT NOT IN THE BRAIN.
-!- DEVELOPMENTAL STAGE: WEAKLY EXPRESSED THROUGHOUT THE EMBRYO,
BUT MARKEDLY UPREGULATED AT DISCRETE SITES DURING DEVELOPMENT. AT
E6.5, PROMINENT EXPRESSION OBSERVED IN DIFFERENTIATED DECIDUA. AT
E7.5, INTENSE EXPRESSION IN EXTRAEMBRYONIC ENDODERM, AMNION AND
NASCENT MESODERM. AT 8.5, ABUNDANT EXPRESSION IN SOMITES AND YOLK
SAC FOLLOWED BY A CONFINEMENT TO DERMATOTOME COMPARTMENT. BETWEEN
E9.5 AND E11.5, ABUNDANT EXPRESSION IN AER (THICKENED ECTODERMAL
CELLS OF LIMB BUDS). AT E12.5, EXPRESSION IN THE LIMBS IS CONFINED
TO THE CONDENSING MESENCHYM SURROUNDING THE CARTILAGE. AT THIS
STAGE, STRONG EXPRESSION ALSO DETECTED IN VERTEBRAL AND FACIAL
CARTILAGE PRIMORDIA AND IN THE MUSCLE OF THE TONGUE. AT E16.5,
ABUNDANT EXPRESSION IN EPITHELIAL CELLS OF THE INTESTINAL VILLI.
ISOFORM A IS MOST ABUNDANT AT ALL STAGES BUT SIGNIFICANT LEVELS OF
ISOFORM B OCCUR AT E12.5.
-!- DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE
ASSISTING THE FOLDING OF THE ZMOGEN WITHIN THE ENDOPLASMIC
RETICULUM.
-!- DOMAIN: AC 1 AND AC 2 (CLUSTERS OF ACIDIC AMINO ACIDS) CONTAIN
SORTING INFORMATION. AC 1 DIRECTS TGN LOCALIZATION AND INTERACTS
WITH THE TGN SORTING PROTEIN PACS-1.
-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
-!- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

EMBL; D17583; BAA04507.1; .
EMBL; D12619; BAA02143.1; .
EMBL; L14932; AAA74636.1; .
PIR; JX0248; JX0248.
PIR; A48225; A48225.
HSP; Q99405; IMPT.
DR MGD; S08.076; .
DR MGD; MGI:97515; Pcsk5.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002884; P_domain.
DR InterPro; IPR000209; Peptidase_S8.
Pfam; PF00082; Peptidase_S8; 1.
Pfam; PF01483; P; PARTIAL.

DR PRINTS: PR00723: SUBTILISIN.
DR ProDom; PD000717; P_domain; 1.
DR SMART; SM00181; EGF; 3.
DR SMART; SM00001; EGF_Like; 2.
DR SMART; SM00261; FU; 22.
DR PROSITE; PS00136; SUBTILASE ASP; 1.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE SER; 1.
KW hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;
KW Cleavage on pair of basic residues; Repeat; Alternative splicing;
KW Transmembrane.
FT SIGNAL 1 34
FT PROPEP 35 116
FT CHAIN 117 1877
FT
FT PROTEIN CONVERTASE SUBTILISIN/KEXIN
FT TYPE 5.
FT DOMAIN 117 1768
FT TRANSMEM 1769 1789
FT POTENTIAL.
FT DOMAIN 1790 1877
FT CYTOPLASMIC (POTENTIAL).
FT DOMAIN 117 452
FT CATALYTIC.
FT DOMAIN 464 602
FT HOMO B.
FT DOMAIN 638 1753
FT CYS-RICH MOTIF (CRM) REGION.
FT DOMAIN 1825 1844
FT AC 1.
FT DOMAIN 1856 1877
FT AC 2.
FT SITE 116 117
FT SITE 521 523
FT CLEAVAGE (AUTO-) (BY SIMILARITY).
FT ACT_SITE 173 173
FT CELL ATTACHMENT SITE (POTENTIAL).
FT ACT_SITE 214 214
FT CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 368 388
FT CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 227 227
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 383 383
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 667 667
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 754 754
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 804 804
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 854 854
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 951 951
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1016 1016
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1220 1220
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1317 1317
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1523 1523
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1711 1711
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1733 1733
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 878 915
FT
FT GEYIDDOGHCTCEASCAKCGPTQEDCISCPTRVLDD
FT ATEESWARGFCMLVKNNKLNQKRVLOQLCCCTCTCTFQ
FT (IN ISOFORM PC5A).
FT VARSPLIC 916 1877
FT MISSING (IN ISOFORM PC5A).
FT SEQUENCE 1877 AA; 209287 MW; EC850E2DF20EAIC3 CRC64;
SQ
Query Match 4.5%; Score 242; DB 1; Length 1877;
Best Local Similarity 19.1%; Pred. No. 1.2e-09;
Matches 221; Conservative 97; Mismatches 382; Indels 456; Gaps 64;
QY 61 COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSLPDPVGRKRC-----TFSCA 108
DB 676 CPPGHYHADKKRC-----RKAPN-CESCFG-----SHGQCLSKCYGYFLNEETSSCV 723
QY 109 S-----GEYLEMKNVCKSGCEGTYSGLSGIKFDEMDLPAGES-----NATFMDTVVG 158
DB 724 TQCPDGSYEDIKNKVCCKSENKAC---IGFHNTCEKGLSLQSGRCSTVEDCQFFN 780
QY 159 PSDSRP-----DGCNNSWIPRGNVIESNR--DDCTVSLIYAVHLKSGYVF 203
DB 781 GHDCQPCHRFCATCSGAGADGCIINCT---EGYVMEGRGVQSCSVS--YVLDHSSEGGYK- 835
QY 204 FEYQVVDNNIFF-----EFFFQNDQCQEMDTT 231
DB 836 -SKRCNDSCLTGNGPGFKNCSSPCSGVLLDGLTCQMGAIKDGKEYIDDQGHQCTCEASC 894
QY 232 DKWVKLTDNGEWGSHVLMKSGTNIYWRITGILMGSKAVKPVLVKNTIEGVAYTSBCF 291
DB 895 AKC-----WG-----PTQEDCISCPVTRVLDDGRC-----VMNCPSWKFEKKOCH 935
QY 292. PCKPCTFSNKKGSCFNQCVCPRTNYSEKAKCEICRKDD-----SQFSGSSECTERPPCTTK 347
FT

```

Db 936 PC-----HYTCQCC-----QSGSPNCTSCRADKHGQERFLYHGECLEN--CPVG 978
QY 348 DY-FQIHT--PCDEEGKTOIMKWIPEKICREDLTDAILRPS-----GE 389
Db 979 HYPAGKHTCLPCPD--NCELCY---NPHICRSMGSIYIIPNHTCQKLECRQGEFQDSE 1033
QY 390 KDCPPCPGFFY---NNGSSCHPCPPGTFSDGTEKRCPCPAGTEPALGFYKXKWNVLPG 446
Db 1034 YBECMPCEGCLGTEDDPACTSCATGYMFERHCYKACPEKT-----FGVKWECRACG 1088
QY 447 NMTSCFNVGNKSCDMGWEVAGDHQISGAGSDNDYLILNLHPGFPPTSMTCATGS 506
Db 1089 ---TNGSCDQHEC-----YWCDEGFLSGSCVQDCG-----PGFH-----GQ 1125
QY 507 ELGRITFVETLCSADCVLYFMVDINRKT--NVVESWGTREKQAYTHIIFKNATFTFT 564
Db 1126 ELGEC-----KCHRACTCTSGYNQCSCEGLQHWGTLWSTWPOVEKGD----- 1174
QY 565 WAFQRTNQG-----ODNRRFINDVMKIYSITATNAVDCVASSCRACALGSEQ 611
Db 1175 W-----NEAVPTKPSLVRSLLODRKW---KVQI-KRDATSONQPCSHSCKTC----- 1219
QY 612 SGSSCPVCPGHHYI-----EKETNOCKEC-----PPD 638
Db 1220 NGLCASCPTGMYLWLQACVPCPOGTWPSVTSGCEKCEDCVSCGADLQCCQLSQPD 1279
QY 639 TVLSIHQ-----VYGKE-ACIPCGPGSKNNO-----DHSVCY--- 669
Db 1280 NTLHLHGRCHYSCPGFAYKDGVCHECSPKTCGNATSCNCEGDFVLHGVCWKTC 1339
QY 670 -----SDCFYHEK-----ENQILHYDFSNSLVGSL----- 696
Db 1340 PEKHAVEGCVKHCPCRCQDCI--HEKTCCKMPDFFLYNDMCHRSCPKSFYPMRQCV 1397
QY 697 -----MNGPSFTSGTYFFHFNLSLCGHEGKKM--ALCTNNITDFTVKE----- 739
Db 1398 CHKNLECGPKED-----DCKVCADTSKALHNLGLCDECPGYKKEENDECRD 1447
QY 740 -----IVAGSDDYTNLV-----GAFVCQSTIIPSESKGFR 769
Db 1448 CPESCLICSSANTCLACRGFTVVDVCTAPKECAAVEWDEGSHRCQ-----PCHKKCSR 1503
QY 770 AALSQSIIADTFTGVVETILKNIN-IKEDMFPVPTSIQIDVHFFYKSTATTSC----- 825
Db 1504 CSGPSE-----DQCYTCPRETFLNNTTCVKE-----CPEGYHTDKDSQOVCVLCHSS 1549
QY 826 ---INGRSTAVKMRNPTKSGAGVISVPSKC-----PACTCDGC----- 861
Db 1550 CRTCEGPHSMQCLSCR-----GWFLGKECLLQCRDGYGYGESTGRCCKDKCKSKSCRG 1604
QY 862 -----TFYFLWESAEC-PLCTEHDF-----HELEGAC--KRGFOETLY 897
Db 1605 PRPTQSCDFTFFLLRSKGQCHRACPHYADQHAQTCERCHPTCDKCSGKEANSCLSC 1664
QY 898 VNNEPKWIKGISLPE 913
Db 1665 VWS--YHLKGICPE 1678

RESULT 2
PKK5_BRACL
ID PKK5_BRACL STANDARD; PRT; 1696 AA.
AC Q9NJ15; Q9NJ16; Q9NJ14;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE (Protein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)
DE (Protein convertase PC6-like) (apc6).
GN PC6.
OS Branchiostoma californiensis (California lancelet) (Amphioxus).
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC Branchiostoma
OX NCBI_Taxid=7738;

```

```

RN
RP SEQUENCE FROM N.A. (ISOFORMS A: B AND C).
RX MEDLINE-20175281; PubMed-10708868;
RA Oliva A.A. Jr., Chan S.J., Steiner D.F.;
RT "Evolution of the prohormone convertases: identification of a
RL homologue of PC6 in the protochordate amphioxus.";
CC Biochim. Biophys. Acta 1477:338-348(2000).
CC
CC -!- FUNCTION: LIKELY TO REPRESENT A WIDESPREAD ENDOPEPTIDASE ACTIVITY
CC WITHIN THE CONSTITUTIVE AND REGULATED SECRETORY PATHWAY. CAPABLE
CC OF CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR
CC PROPEPTIDES BY CLEAVAGE OF ARG-XAA-YAA-ARG-I-ZAA BONDS, WHERE XAA
CC CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.
CC -!- SUBCELLULAR LOCATION: ISOFORM A AND ISOFORM C ARE SECRETED.
CC ISOFORM B IS A TYPE I MEMBRANE PROTEIN.
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; A, B (SHOWN HERE) AND C: ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE
CC ASSISTING THE FOLDING OF THE ZYMOGEN WITHIN THE ENDOPLASMIC
CC RETICULUM.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
CC -!- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF184615; AAF26300.1; -
CC EMBL; AF184616; AAF26301.1; -
CC EMBL; AF184617; AAF26302.1; -
CC HSSP; Q99405; IMPT.
CC
CC MEROPS; S08.UPB; -.
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR002884; P_domain
CC InterPro; IPR000209; Peptidase_S8.
CC Pfam; PF00082; Peptidase_S8; 1.
CC Pfam; PF01483; P; 1.
CC PRINTS; PR00723; SUBTILISIN.
CC ProDom; PD000717; P_domain; 1.
CC SMART; SM00181; EGF; 2.
CC SMART; SM00261; FU; 17.
CC PROSITE; PS00136; SUBTILASE_ASP; FALSE_NEG.
CC PROSITE; PS00137; SUBTILASE_HIS; 1.
CC PROSITE; PS00138; SUBTILASE_SER; 1.
CC Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;
CC Cleavage on pair of basic residues; Repeat; Alternative splicing;
KW Transmembrane. 1 25 POTENTIAL.
KW SIGNAL. 1 25 POTENTIAL.
FT PROPEP 26 110 POTENTIAL.
FT CHAIN 111 1696 PROPROTEIN CONVERTASE SUBTILISIN/KEXIN
FT TYPE 5.
FT
FT DOMAIN 111 1618 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1619 1639 POTENTIAL.
FT DOMAIN 1640 1696 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 111 488 CATALYTIC.
FT DOMAIN 496 637 HOMO B.
FT DOMAIN 664 1649 CYS-RICH MOTIF (CRM) REGION.
FT SITE 110 111 CLEAVAGE (AUTO-) (BY SIMILARITY).
FT ACT_SITE 192 192 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 233 233 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 407 407 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 246 246 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 529 529 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 885 885 N-LINKED (GLCNAC...) (POTENTIAL).
FT VARSPIC 1259 1323 DDTILDRGECITSCGPGYMDREKKCKACHPTCKECSDEY
FT DDTCTACNDGFLTLTDPASSCEAGCP -> AENQNASFCPFA
FT PREVSVALAEIHLRLYSLTDPQSNPPDPTVLGADRRL
FT TTATSAAGRCA (IN ISOFORM C).
FT

```

```
FT VARSPLIC 1324 1696 MISSING (IN ISOFORM C).
FT CHPTCKECSDEYDTCACNDGFLTLTASSCEAGCPGQFL
FT HHGDCDSCHREKTC -> IARCVDRDRSNCIDLVRNFC
FT VRRYEVKRCCTCKLYMEDRPMRRGSSQPTQGRN (IN
FT ISOFORM A).
FT VARSPLIC 1344 1696 MISSING (IN ISOFORM A).
FT SEQUENCE 1696 AA; 188410 MW; 281CBE1784257CBD CRC64;

Query Match
Best Local Similarity 4.5%; Score 240; DB 1; Length 1696;
Matches 206; Conservative 98; Mismatches 358; Indels 414; Gaps 57;

QY 35 CCWALAGQAQWAGDLPSSSRPLPCQKDYHFE-----YTECDSSGSRWRVAIPNSAV 89
DB 680 CHDSACTCHRGSGCGLS-----CHGKNFVEDEGTCEECGQ--GYKDEBERKCL 729

QY 90 DCSG--LPDPVKGECTFCSCASGEYLEMKNQVC-SKCGEGTYSLSGGIKFDEWDELPAGF 146
DB 730 DCSADCLTCQVADHCT-SCDDEGLKLFENTCVAQCSEGRY-----MDENDVC--- 777

QY 147 SNLAFMDTVVGPDSRSDPCNNSWIPRGNTIESNRDDCTVSLIYVHLKSGGVYFEY 206
DB 778 QDCDDSCDCTGPDATDCVTCADDELLQESQVES-----CS-----SGYFOQY 822

QY 207 QYVDNNIFFEFP--IQNDQC-----QEMDTTDDKWKLTDNGEMSH-----SVMLK 251
DB 823 ECLKCHATCASCGRSDDOCLTCSGHLLEDHTRCITSCDEGEYTERGKCEDCNICK 882

QY 252 S-----GTNLYWRTTGLMGSKAVRPLVKNITIEGVATVSCFPCKPOT 297
DB 883 KNGSQADQCLCHDHTN-LYDTCVOYCGNRRYPE-----NGECHCPHPS 928

QY 298 FSNKPGSFN-----CQV--CPRNTYSEKG-----AKECIRCKD 329
DB 929 LCCIGEIQCITDYEGEDHFLYQGTCHVTCPLGLYDITDQVCKACAPCTIACDGP 988

QY 330 SFGSSSECT---ERPCTTKDYFOIHPDCEEGKQIMYKWE-----PKI---C 374
DB 989 A---DNOCITCEERAPTDR-----CQSEGSQTDABCAEGCHSCERGPDI CDSC 1036

QY 375 RED--LTD--AIRLP-----PSGKKDCPPCN-----PGFYNNCS 405
DB 1037 DEYLYTDTCTVRRNCPSTFTPDODRRCRPHCHDCEACDGNQNCSCKEGFKYTPD 1096

QY 406 SSCHPCPPTFSDGT--KECRCPAGTEPALGEYKWNVLPGNMKTSCFNVGNSCKDGMN 464
DB 1097 GCSTGCPNRYKDDTNKECKPC-----DSSCFT-----CSG-- 1127

QY 465 GNEVAGDHQSGAGSDNDYLILNLHIPGFKPPTSMGTATGSELGRITVFTVFLCSADCV 524
DB 1128 ---PASFHCLSCADGD-----FLHSSCRSTCP 1152

QY 525 LYFMVDINKSTNVVBSWGTGTEKQAYTHIIFKNATFTFTWAFQRTNOGQ-----D 575
DB 1153 AGFIG--NAESHECVS--SCQDQYYS-----SETGRCEDCPYNCRAD 1193

QY 576 NRRFINDMKIYSI-----TATNAVDDGVASSACALGSEQSG 613
DB 1194 NDGDCAECAPTIVVDRCRPEETCEDGEYQDRDRDTAELSRCPHOSCKTCSGSPSDTDC 1253

QY 614 SSC-----VPCPPGHYTEKETNOCKECP-----DYLSLTHQVYKREACTP 654
DB 1254 DSKGDDTILDRGECITSCGPGGEYMDRREKCKKACHPTCKECSDEY-----DDTCTA 1305

QY 655 CPGSKNNODHSVCYSDCFYHEKENQIILHY--DFSLLSSVSGSLMNGPSTSKGTYFHF 713
DB 1306 CNDGFLLT--DASSCAGC-----PPQFLHGDGDCSCHRECKTCGDP-----HID 1349

QY 714 NISLCHCEGKMK--ALCTNITDFTVKEIVAGSDDYTNLGVAFVQSTIIPSESGFRAA 771
DB 1350 NCLSC--QPGSYLNDQOCSTHCPGEGTPEE-----TYEDDSGETVLQCLCHVNCK----- 1397

QY 772 LSSQSILADTIGTVTETLTK---NINKEDMFPVPTSQIPDVHFFYKSSATTSCING 828
```

```
DB 1398 -----TCHGEEDCEMCANDIKYKQD-----GRCVTECOEG 1429
QY 829 RSTAVKMRNPTKSGAGVISVPSKCPAGCTCDGCTFFYFLWESAECPLCTEHDHFELEGA- 887
DB 1430 HYPDLTNECQCWS-----DCETCDG-----PRNDQCVC-PYNYLVLGKC 1470
QY 888 ---CRRGFQETLYVWNEPKWCIKIGISLPEKKLATCE-----TVDFWLKVGAGV 932
DB 1471 LEDCPGEGYDTHWQKECECEC-----HPSCATCNEGNGYNCLSCPYGSKLGEV 1519

RESULT 3
VS41_GIALA
ID VS41_GIALA STANDARD; PRT; 687 AA.
AC P92127;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Variant-specific surface protein VSP4A1 precursor (CRISP-90).
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardinales; Giardia.
OX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=O2-4A1;
RX MEDLINE=97321554; PubMed=9178264;
RA Papanastasiou P., Bruderer T., Li Y., Bommeli C., Koehler P.;
RT "Primary structure and biochemical properties of a variant-specific
surface protein of Giardia.";
RL Mol. Biochem. Parasitol. 86:13-27(1997).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=97233006; PubMed=9078242;
RA Papanastasiou P., McConville M.J., Ralton J., Koehler P.;
RT "The variant-specific surface protein of Giardia, VSP4A1, is a
glycosylated and palmitoylated protein.";
RL Biochem. J. 322:49-56(1997).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ANCHORED TO THE
PLASMA MEMBRANE.
CC -!- PTM: O-GLYCOSYLATED. THE MAJOR GLYCAN IS A TRISACCHARIDE WITH GLC
AT THE REDUCING TERMINUS.
CC -!- PTM: PALMITOYLATED.
CC -!- SIMILARITY: BELONGS TO THE GIARDIA VARIANT SURFACE PROTEIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; 283743; CAB06038.1; .
CC HISSP; O14763; IDUG.
CC GlycoSuiteDB; P92127; .
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR005127; Giardia_vsp.
CC Pfam; PF03302; VSP; 2.
CC SMART; SM00181; EGF; 2.
CC SMART; SM00001; EGF_like; 1.
CC SMART; SM00261; FU; 3.
CC Antigen; Repeat; Transmembrane; Glycoprotein; Lipoprotein; Palmitate;
Signal.
CC SIGNAL. 1 14
CC CHAIN 15 687
CC DOMAIN 15 660
CC TRANSMEM 661 681
CC DOMAIN 682 687
CC CYTOPLASMIC (POTENTIAL).
CC SEQUENCE 687 AA; 70857 MW; D892F675D626D7EC CRC64;

Query Match
4.2%; Score 223.5; DB 1; Length 687;
```

Best Local Similarity 18.7%; Pred. No. 7.9e-09;
Matches 144; Conservative 88; Mismatches 255; Indels 283; Gaps 36;

QY 290 CFPCKPCTF-----SNKPGSFNCVCPRTNYSEKAKEIRCKDDSQSGSSECTE 340
Db 83 CGKCGDGYFLMGCGYKTESQPGS---EIC---TTASNGL--CTACKVDSQY----- 126
QY 341 RPPCTTKDYFIHTPCDEEGTKIIMYKWIPEKICREDLTDALRLPPSGEKKDCPPCNPGE 400
Db 127 -----IFONKATPSEKSECLC-W-----DSTD--RNGVMG--VANCATCTAPA 166
QY 401 YNNGSSCHPCPGTSDGTEKCRPCAGTEPALGFEYKWNVLPNMTKTSFCFNVGNSKC 460
Db 167 SSTGPATCTECMAGTYKKSDTECAACHSDC-----ATCSGEANNQC 207
QY 461 DGMNGWEVAGDHQSGAGGSDNDYLILNHPGFKPPTSWT--GATGSELGRITFFFEI- 517
Db 208 TSCE---TGKYLKSNQCEKN---TCNTHYPDDTSMTCVACTVLDANCACTCFDSA 258
QY 518 -----LCSADCVLFYMDINRKSTNVVESWGGTKKQAYTHIIFKNATFTTWFQRT 570
Db 259 TAKGKCLTCSNKNKIPRTLD--GTSTCVENSYAGC----- 291
QY 571 NQODNRIND---WKIYSITATNWDGVASSCRACALGSEQSGSSVCPPPGHYIEK 627
Db 292 -QGADNELFMKEDQSACLLCGDGTKEASNDKGVANCRTCTKNANDSPPTCTACLDGYFLER 350
QY 628 ET-----NCKECPDPTILSIHOVYCKEACIPC----- 655
Db 351 GSCITTCADNATCSEATTEDKCKICKAGFFLASP---GEGKICSDTNNGIDGCAEC 407
QY 656 -----GSKNN-----QHSVCYSDCFFHYHEKENQILHYDFSN 689
Db 408 TKPAGPLKCTCKPNRKPAGTSDNYTEKTCEDPTVC----- 446
QY 690 LSSVGLMNGPFTSGTKTYFFHFNLSLGHGKKM--ALCTNN-----ITDFTVKEIV 741
Db 447 -GCTSACDAIVDANGKEHY---CSYGETNKFPIDGLCTDNKGTAGCTDHTCSYCA 502
QY 742 AGSDDY-----TNLVGAFVQOST-----IPSESKGF-----RAALSSQSILLADTFI 784
Db 503 AGFFLYMGCKYKIDTVPGSYMCKSKSTAGVCDTPNANNRFFVVPKRAISEQSVLAGCNPL 562
QY 785 GTVETTLNINIKEDMFVPTSQIPDVHFFYKSSAT--TSCINGRSTAVKMRCPNPKSG 843
Db 563 G-----TIAGGNAYGVGECSCQTAAPDARADGMAMAVTCTACEDGK-----KPKKSG 609
QY 844 AGVISVPSKCPAGTCDCGCTFYFLWESAECPLCTEHDHFIEGACKRGQETLYVMNEPK 903
Db 610 TCGVA---CPDANKSCIT-----MDDVCEE----- 631
QY 904 WCIGKISLPEKKLATCETVDFWLKVGAGVGAFTAVL-----LVALTCYFW 948
Db 632 -CADGFSLDNGKCVSSGNTKNSGLSTGAINGISVAALVVVGGVLGLCWF 680

RESULT 4
TSA4-GIALA
ID TSA4-GIALA STANDARD; PRT; 713 AA.
AC P21849;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Major surface-labeled trophozoite antigen 417 precursor.
GN TSA 417.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
OX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 30957 / WB;
RX MEDLINE=90280395; PubMed=2352929;
RA Gillin F.D., Hagblom P., Harwood J., Aley S.B., Reiner D.S.,

McCaffery M., So M., Guiney D.G.;
"Isolation and expression of the gene for a major surface protein of
Giardia lamblia";
Proc. Natl. Acad. Sci. U.S.A. 87:4463-4467(1990).
[2]
SEQUENCE OF 480-620 FROM N.A.
RC STRAIN=AD-1;
RX MEDLINE=93314970; PubMed=8325510;
RA Ey P.L., Mayrhofer G.;
"Two genes encoding homologous 70-kDa surface proteins are present
within individual trophozoites of the binucleate protozoan parasite
Giardia intestinalis";
Gene 129:257-262(1993).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ANCHORED TO THE
PLASMA MEMBRANE.
CC -!- DOMAIN: CONTAINS 29 REPEATS OF THE CXXC MOTIF.
CC -!- SIMILARITY: BELONGS TO THE GIARDIA VARIANT SURFACE PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC
CC EMBL; M33641; AAA02688.1; -;
DR EMBL; M97488; AAA02581.1; -;
DR PIR; A35502; A35502.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR005127; Giardia_VSP.
DR Pfam; PF03302; VSP; 2.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00001; EGF_Like; 1.
DR SMART; SM00261; FU; 3.
KW Signal; Antigen; Glycoprotein; Transmembrane; Repeat.
FT SIGNAL 1 17
FT CHAIN 18 713 MAJOR SURFACE-LABELLED TROPHOZOITE
FT DOMAIN 18 679 ANTIGEN 417.
FT TRANSMEM 680 708 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 709 713 POTENTIAL.
FT CARBOHYD 289 713 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 676 676 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 582 582 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 606 606 A -> T (IN STRAIN ADELAIDE-1).
FT VARIANT 606 606 A -> S (IN STRAIN ADELAIDE-1).
SQ SEQUENCE 713 AA: 72510 MW: 9AD7195843DE5601 CRC64;
Query Match 3.9%; Score 210; DB 1; Length 713;
Best Local Similarity 19.7%; Pred. No. 8.2e-08;
Matches 186; Conservative 96; Mismatches 300; Indels 360; Gaps 53;
QY 90 DCSGLP-DPVRGKECTFCASGEYLEMKNQ----VCSKCG-----EGTYSILGSGIK-- 135
Db 43 ECGNANYAPVNG-QCVDVNAEGPSKTLCPQHSAGKCTCGGNSFMYKDCYSSGEGLPCH 101
QY 136 -----FDEWDEL-----PAGFSN-----INTFMDTVVGPSPD-----S 162
Db 102 SLCLSSDGDGVCTEAAPGYFAPVGAANTEQSVIACGDTTGTVTIAGGNTYKGIADCAECS 161
QY 163 RPDGCGNNSWIPIRGVNIENRDDCTVSLIYAVHLKKGVSFFVEYQVVDNNIFFEIQND 222
Db 162 AFDA---TAGAAGKVAUTC--CGVS-----KYLKDNVCV-----D 193
QY 223 QCQEMDTTDDKWKVLTNDGEMGSHSVMLKSGTNYLWRTTGLMGSKAVKPVLVKNITIE 282
Db 194 KAQCNSGSGTNKEVAV--DDESGNKVCSCSDNLN-----G 226
QY 283 GVAYTSECFPCPGTFENKPGSFNCGVCPNRYTSKGAKE-ECIRCKDDSQSGSSE---C 338
Db 227 GVA-----NCDTC---SYDQSKKIKCTKCTDNNYLTSTGTSCT 263

```
QY 339 TERPPCTTKDYIHTPCDEGKTOIMYKWIIEPKICREDLTDAILRPPSGEKKDCPPCNP 398
DB 264 VKQDQC--KDFG---FPRKDDSS-----AGNK--CLPCND 290
QY 399 GRYNNGSSSCHPCP--PGTFSOGTEKPCPAGTBPALGFYKWNVLPGNMKWTSCFNVG 456
DB 291 S--TDCIANCATCALVSRGAALVTCSSACTDGYKPSAD-----KTTCEAVS 335
QY 457 NSKCDGNGWVAGDHQISGAGSDNDYLILNLHIPKFPPTSMTCATSELGRITTFVFE 516
DB 336 NCKTPCKKACNEGENEVCTDCDSTYL-----TPFSQ----- 369
QY 517 TLCSADVLVFMVDINRKSTNVESGGTKEQAVTHIIFKNATFTFWAFORT--NQCQ 574
DB 370 --CIDSCA-----KIGNYGADEGAKK-----LCKECTA-----ANCKTCDQDQ 407
QY 575 DNRREFNDMWKIYSTATNAVGVASSRACALGSEQSGSSVCPGPGHYIEK-----ET 629
DB 408 -----COACNDGFYKNGDAGSPC---HESCKTCSAGTA 437
QY 630 NOCKECPDPTILSIHOVYGEACI--PCGPGSKNNQDHSVCYSDCFFYHEKENQILHYDFS 688
DB 438 SDCTECPTKALR----YDDGTGTCGEGCTGTGAGAC----- 473
QY 689 NLSSVGLMNGPSTSKGTYHFFFNISLCGHEGKKMALCTNNITDFTVKEIVAG--SD 746
DB 474 --KTCGLTIDGASYCECATTTTEYPQNGVC---APKASRATPTCNDSPIONGVCGTCADN 528
QY 747 YNVLGCAVFCQSTIIPSESKGFRAALSSOSIILADTFIGVTVELTKNINIKEDFPVPT 806
DB 529 YFKMNGG--CYETV--KYPGKTVICISAPN-----GGTCQKAADGYKLDGSLTVCS 575
QY 807 SQIPDVHFFYKSTATTSCING--RSTAVKMRCN---PTKSGAGVISVSPKCPAGTCDGC 861
DB 576 EGCKEC-----ASTDCTCLDGVYKVSASACTKODASCETCNGA-----ATTCKAC 621
QY 862 -TFYFWESAE-ACPLCTEHDHEIEGACKRGFQETLYVYNNPKWICIGISPEKK-LAT 918
DB 622 ATGYKYKTASGEGACTSC-ESDSNGVTGI-----KGLNCAPPPNNKRGSVL 665
QY 919 C-----ETVDPLKVGAGVAGTAFL-----LVALTCTYFW 948
DB 666 CYLIKDSGSTNKSGLTGAIGASIVAVVVVGLGIFLGCWF 707

RESULT 5
LMA5_MOUSE
ID LMA5_MOUSE STANDARD; PRT: 3718 AA.
AC Q61001; Q9JHQ6;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Laminin alpha-5 chain precursor.
GN LMA5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE OF 1-92 FROM N.A.
RA Timpi R., Sasaki T.;
RT "Completion of the N-terminal sequence of the murine Laminin alpha 5
chain.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 84-3718 FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Lung;
RX MEDLINE=96081906; PubMed=7499364;
RA Miner J.H., Lewis R.M., Sanes J.R.;
RT "Molecular cloning of a novel laminin chain, alpha 5, and widespread
expression in adult mouse tissues.";
RL J. Biol. Chem. 270:28523-28526(1995).
RN [3]
```

```
RP REVISIONS.
RA Miner J.H., Lewis R.M., Sanes J.R.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Binding to cells via a high affinity receptor. Laminin
is thought to mediate the attachment, migration, and organization
of cells into tissues during embryonic development by interacting
with other extracellular matrix components.
CC -!- FUNCTION: ALPHA-5 CHAIN MAY BE THE MAJOR LAMININ ALPHA CHAIN OF
ADULT EPITHELIAL AND/OR ENDOTHELIAL BASAL LAMINAE.
CC -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
different polypeptide chains (alpha, beta, gamma), which are bound
to each other by disulfide bonds into a cross-shaped molecule
comprising one long and three short arms with globules at each
end.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT
MEMBRANES (MAJOR COMPONENT).
CC -!- TISSUE SPECIFICITY: IN ADULT, HIGH LEVELS IN HEART, LUNG, AND
KIDNEY; LOWER IN BRAIN, MUSCLE AND TESTIS; VERY LOW IN LIVER, GUT
AND SKIN. EXPRESSED IN MANY TISSUES IN EMBRYONIC DAY 11.
CC -!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC -!- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.
CC -!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -!- SIMILARITY: CONTAINS 21.5 LAMININ EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.
CC -!- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ293593; CAB99255.1; --
CC EMBL; U37501; AAC53430.1; --
CC HSP; P02468; ITLE.
CC MGD; MGI:105382; Lama5.
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR001886; LamNT.
CC InterPro; IPR000034; Laminin_B.
CC InterPro; IPR002049; Laminin_EGF.
CC InterPro; IPR001791; Laminin_G.
CC Pfam; PF00052; laminin_B; 1.
CC Pfam; PF00053; laminin_EGF; 19.
CC Pfam; PF00054; laminin_G; 2.
CC Pfam; PF00055; laminin_Nterm; 1.
CC ProDom; PD002082; LamNT; 1.
CC ProDom; PD003031; Laminin_B; 1.
CC PROSITE; PS00022; EGF_1; 19.
CC PROSITE; PS01186; EGF_2; 3.
CC PROSITE; PS01248; LAMININ_TYPE_EGF; 19.
CC PROSITE; PS00025; LAM_G_DOMAIN; 5.
KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
FT SIGNAL. 1 40
FT CHAIN. 41 3718
FT DOMAIN 41 304 LAMININ ALPHA-5 CHAIN.
FT DOMAIN 305 363 LAMININ N-TERMINAL (DOMAIN VI).
FT DOMAIN 364 433 LAMININ EGF-LIKE 1.
FT DOMAIN 434 479 LAMININ EGF-LIKE 2.
FT DOMAIN 500 546 LAMININ EGF-LIKE 3.
FT DOMAIN 547 592 LAMININ EGF-LIKE 4.
FT DOMAIN 593 637 LAMININ EGF-LIKE 5.
FT DOMAIN 638 682 LAMININ EGF-LIKE 6.
FT DOMAIN 683 728 LAMININ EGF-LIKE 7.
FT DOMAIN 729 781 LAMININ EGF-LIKE 8.
FT DOMAIN 782 833 LAMININ EGF-LIKE 9.
FT DOMAIN 834 855 LAMININ EGF-LIKE 10.
FT DOMAIN 856 1442 LAMININ EGF-LIKE 11 (INCOMPLETE).
FT DOMAIN 1443 1488 LAMININ DOMAIN IV 1 (DOMAIN IV B).
FT DOMAIN 1489 1532 LAMININ EGF-LIKE 12.
FT DOMAIN 1499 1532 LAMININ EGF-LIKE 13.
```

FT	DOMAIN	1533	1581	LAMININ EGF-LIKE 14.
FT	DOMAIN	1582	1632	LAMININ EGF-LIKE 15.
FT	DOMAIN	1633	1642	LAMININ EGF-LIKE 16 (N-TERMINAL).
FT	DOMAIN	1643	1831	LAMININ DOMAIN IV 2 (DOMAIN IV A)
FT	DOMAIN	1832	1864	LAMININ EGF-LIKE 16 (C-TERMINAL).
FT	DOMAIN	1865	1914	LAMININ EGF-LIKE 17.
FT	DOMAIN	1915	1970	LAMININ EGF-LIKE 18.
FT	DOMAIN	1971	2024	LAMININ EGF-LIKE 19.
FT	DOMAIN	2025	2071	LAMININ EGF-LIKE 20.
FT	DOMAIN	2072	2118	LAMININ EGF-LIKE 21.
FT	DOMAIN	2119	2168	LAMININ EGF-LIKE 22.
FT	DOMAIN	2169	2735	DOMAIN II AND I.
FT	DOMAIN	2736	2933	LAMININ G-LIKE 1.
FT	DOMAIN	2947	3119	LAMININ G-LIKE 2.
FT	DOMAIN	3128	3296	LAMININ G-LIKE 3.
FT	DOMAIN	3337	3511	LAMININ G-LIKE 4.
FT	DOMAIN	3518	3689	LAMININ G-LIKE 5.
FT	DOMAIN	2205	2257	COILED COIL (POTENTIAL).
FT	DOMAIN	2330	2464	COILED COIL (POTENTIAL).
FT	DOMAIN	2604	2621	COILED COIL (POTENTIAL).
FT	DOMAIN	2639	2705	COILED COIL (POTENTIAL).
FT	SITE	1723	1725	CELL ATTACHMENT SITE (POTENTIAL).
FT	SITE	1839	1841	CELL ATTACHMENT SITE (POTENTIAL).
FT	DISULFID	305	314	BY SIMILARITY.
FT	DISULFID	307	327	BY SIMILARITY.
FT	DISULFID	329	338	BY SIMILARITY.
FT	DISULFID	341	361	BY SIMILARITY.
FT	DISULFID	364	373	BY SIMILARITY.
FT	DISULFID	366	398	BY SIMILARITY.
FT	DISULFID	401	410	BY SIMILARITY.
FT	DISULFID	413	431	BY SIMILARITY.
FT	DISULFID	434	445	BY SIMILARITY.
FT	DISULFID	436	452	BY SIMILARITY.
FT	DISULFID	454	463	BY SIMILARITY.
FT	DISULFID	466	476	BY SIMILARITY.
FT	DISULFID	500	512	BY SIMILARITY.
FT	DISULFID	502	521	BY SIMILARITY.
FT	DISULFID	523	532	BY SIMILARITY.
FT	DISULFID	535	544	BY SIMILARITY.
FT	DISULFID	547	559	BY SIMILARITY.
FT	DISULFID	549	566	BY SIMILARITY.
FT	DISULFID	568	577	BY SIMILARITY.
FT	DISULFID	580	590	BY SIMILARITY.
FT	DISULFID	593	605	BY SIMILARITY.
FT	DISULFID	595	611	BY SIMILARITY.
FT	DISULFID	613	622	BY SIMILARITY.
FT	DISULFID	625	635	BY SIMILARITY.
FT	DISULFID	638	650	BY SIMILARITY.
FT	DISULFID	640	656	BY SIMILARITY.
FT	DISULFID	658	667	BY SIMILARITY.
FT	DISULFID	670	680	BY SIMILARITY.
FT	DISULFID	683	695	BY SIMILARITY.
FT	DISULFID	685	702	BY SIMILARITY.
FT	DISULFID	704	713	BY SIMILARITY.
FT	DISULFID	716	726	BY SIMILARITY.
FT	DISULFID	1443	1455	BY SIMILARITY.
FT	DISULFID	1445	1462	BY SIMILARITY.
FT	DISULFID	1464	1473	BY SIMILARITY.
FT	DISULFID	1476	1486	BY SIMILARITY.
FT	DISULFID	1533	1548	BY SIMILARITY.
FT	DISULFID	1535	1555	BY SIMILARITY.
FT	DISULFID	1557	1566	BY SIMILARITY.
FT	DISULFID	1569	1579	BY SIMILARITY.
FT	DISULFID	1582	1594	BY SIMILARITY.
FT	DISULFID	1584	1601	BY SIMILARITY.
FT	DISULFID	1603	1612	BY SIMILARITY.
FT	DISULFID	1615	1630	BY SIMILARITY.
FT	DISULFID	1865	1874	BY SIMILARITY.
FT	DISULFID	1867	1881	BY SIMILARITY.
FT	DISULFID	1884	1893	BY SIMILARITY.
FT	DISULFID	1896	1912	BY SIMILARITY.
FT	DISULFID	1915	1930	BY SIMILARITY.
FT	DISULFID	1917	1939	BY SIMILARITY.

FT	DISULFID	1941	1950		BY SIMILARITY.
FT	DISULFID	1953	1968		BY SIMILARITY.
FT	DISULFID	1971	1986		BY SIMILARITY.
FT	DISULFID	1973	1993		BY SIMILARITY.
FT	DISULFID	1996	2005		BY SIMILARITY.
FT	DISULFID	2008	2022		BY SIMILARITY.
FT	DISULFID	2072	2083		BY SIMILARITY.
FT	DISULFID	2074	2090		BY SIMILARITY.
FT	DISULFID	2092	2101		BY SIMILARITY.
FT	DISULFID	2104	2116		BY SIMILARITY.
FT	DISULFID	2119	2126		BY SIMILARITY.
FT	DISULFID	2121	2133		BY SIMILARITY.
FT	DISULFID	2135	2144		BY SIMILARITY.
FT	DISULFID	2147	2166		BY SIMILARITY.
FT	DISULFID	2169	2169		INTERCHAIN (PROBABLE).
FT	DISULFID	2172	2172		INTERCHAIN (PROBABLE).
FT	CARBOHYD	100	100		N-LINKED (GLCNAC. .)
FT	CARBOHYD	148	148		N-LINKED (GLCNAC. .)
FT	CARBOHYD	248	248		N-LINKED (GLCNAC. .)
FT	CARBOHYD	383	383		N-LINKED (GLCNAC. .)
FT	CARBOHYD	457	457		N-LINKED (GLCNAC. .)
FT	CARBOHYD	485	485		N-LINKED (GLCNAC. .)
FT	CARBOHYD	905	905		N-LINKED (GLCNAC. .)
FT	CARBOHYD	926	926		N-LINKED (GLCNAC. .)
FT	CARBOHYD	964	964		N-LINKED (GLCNAC. .)
FT	CARBOHYD	1335	1335		N-LINKED (GLCNAC. .)
FT	CARBOHYD	1534	1534		N-LINKED (GLCNAC. .)
FT	CARBOHYD	2021	2021		N-LINKED (GLCNAC. .)

Query Match 3.7%; Score 200; DB 1; Length 3718;

Best Local Similarity 18.5%; Pred. No. 3.7e-06;

Matches 174; Conservative 82; Mismatches 279; Indels 404; Gaps

QY	6	RGPVRRG-----	--WGRPAEAPRGRSPPWSPANICWMALAGCOAAWAGDLPS-- S 54
Db	1465	RGHVIGRDSCATGYWGFPNCRP-	-----CDGARLCDLTGCQTCP 1506
QY	55	SRLPLP----	COEKDYHE-----YTECDSG-SRWRAIPNSAVD---CSGLPDVPVRKE 102
Db	1507	PRTPPDCLVQPQSGCHPLVGCEECNC	SGVGVELTPTCDMDSDGGRCRPN-VAGR 1565
QY	103	CTFSCASGEY-----	LEMKNQV----CSKCGEGTYSLGS 132
Db	1566	CD-TCAPFGYGVSPCRPCDHEAGTMASVC	DPLTGCHCKENVOGSRCDQCVRGTFSIDA 1624
QY	133	G-----IKDEWDELPAESNTATP-----	MDTVVGPSDRP----- 164
Db	1625	ANPKGTRCFCEGATER--CGSNSLARHEFVDMEGWLLSSDRQVVP	HEHRPETELLHAD 1682
QY	165	----DCGNSSWI	PNGYNIESNRDDCTSLIYAVHLK-KSGYVFEEYOYDNNIPEFF 218
Db	1683	LRSVAOTFSELYWAQPSYGLDRVS	SYGGTGLHYLEHSETQRGDIFIPYESRPDVV----- 1737
QY	219	IQNDOQQ-----	EMDTTDKKWL-TDNGEWSHSVMLKSGTNILYWR 260
Db	1738	LOGNQMSIAFLAYPPGOVHQQLQVLG	VGNFRHLETHNPVSREELMMVLAGLEQLQIR 1797
QY	261	TTOILMGSKAVPVLVNITISGAVT-----	SECFFCKPGTFSN 300
Db	1798	A-----LFSOTSSSVLSLRVVLEVASEA	RGPPASPASNVELCMCPANYRGSDCOCEAPGYRD 1853
QY	301	KPGSF--NCQVCPRTNYSEK-----	GAKECIRCKDDSQFSGSSECTERPPCTTKDYFOIHTP 355
Db	1854	TKGLFLGRCPVCCOCHGSDRCLPGSGIC	VGCOHNTEQDOCERC--RPGFVSDSPNPASP 1911
QY	356	CD-----	BEGKTQIMYKWIEPKICREDLTDAILRPPSGEKKDCPPC 396
Db	1912	CVSCPCPLAVPSNNFADGCVLNGRTQC-----	LCR-----PGYAGASCRC 1953
QY	397	NPGFYNNG---SSSCHPC-----	PPGTFSDG---TKECRPC-----PASTEPALGF 437
Db	1954	APGFEFNPLVSSCQPCDCSGNDGPNMTFSD	CDPLTGACRGCLRHTTGPHCERCAFGF- 2012


```
FT CARBOHYD 499 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1216 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1239 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1314 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1814 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1818 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1908 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1933 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2028 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2111 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2142 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2332 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2533 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2575 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2692 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2812 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3052 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3055 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3144 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3172 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3288 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3292 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3782 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4005 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4136 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4243 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4254 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4335 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4376 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 4586 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 5136 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 5252 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 5376 AA: 579908 MW: 0E44DB77DF2A2620 CRC64;

Query Match 3.6%; Score 194; DB 1; Length 5376;
Best Local Similarity 20.5%; Pred. No. 1.7e-05;
Matches 144; Conservative 63; Mismatches 204; Indels 290; Gaps 38;

QY 75 SSGSRWRVAIPNSAVDCSLDPVRKECTFSCAGSEYLEMKNQVCKGEGTYSLGSGI 134
DB 3256 SRGCTQSCTCPAGATHCR-----NFKCPSTYCKNGDNGSSNCTEITLQCPPTS 3304
QY 135 KFEDEWELPAGESNATFMDTVGVGSDSPDCGNSSWTPRPNYTESNRDDCTVSLIYAV 194
DB 3305 QET--DCLP-----SCVPSCSNRCEVTSPTSPVSSCREG-----3336
QY 195 HLKKGYPVFEXYVDNNIFFEFTQNDQCQEMDTTDTKWKVLTDNGWGSVLMKSGT 254
DB 3337 -LCNHGFVFE-----DKVPTQCGCKDARGAIIIPAG- 3368
QY 255 NIIYWRITGLMGSKAVKPVLYKNITIEGVAVTSECFCKPGTFSNKGPSFNCQ--VCPR 312
DB 3369 -----KTWTSKGTQSCACV-----EGNIOQCNFOCPP 3396
QY 313 NTYSEKGAKECIRCKDDSOFSSESSECTE---RPPCTTKDYFOIHTPC-----DEEG 360
DB 3397 ETY-----CKDNSE--GSSTCTKITLQCPAHTQ-----YTSCLPSCPLSCLDPREG 3439
QY 361 KTOIMYKWIPIKPI---CRE-----DLTDAI-RLPPSGERKKDCPP 395
DB 3440 ----LCKDLSPKVPSTCREGVCQSGYLVNSDKVLRABCDCKDAQAGALIPAGKWTSPG 3495
QY 396 CNPG-FYNNGSSSCH--PCPGPTF-----SDGTKECR-----PCPAGTEPALGFEYKWNVL 444
DB 3496 CTQSCACMGAVQCCSQCPPTCYCKDNEDGNSCAKITLQCPAHS-----LFTNCL 3547
QY 445 PCNMKTSCFNVNSKCDGNGWAGVADHTQSGAGSDNDYLILNLIHPIGPKPTSMGTAT 504
DB 3548 PPCL-PSCLDP-----DGL-----CKGASPKVPSTCKE 3574
QY 505 GSELGRITVFETLCSADCLVLYFMVDINKSTN---VVESWGCTKEQAYTHIIFKNATFT 562
DB 3575 G-----CICQSGVYL-----SNKCLLRNRCCKD--AHGALIPEDK----3609
```

```
QY 563 FTWAFORTNOGDNRFRINDMWKIYSTATNAVDGVASSCRACALGSEQSGSSCVPCPPG 622
DB 3610 -TWWSR-----GCTQSC-VCTGSGIOCLSS--QCPPG 3637
QY 623 HYI---EKETNOCKEGPPDPTLYSIHQVYKGEACIPGPGSKNNQDH-----SVCYS 670
DB 3638 AVCKDNEDGSSNCARIPPOCPANSHYT---DCFPCCPPSCSDPEGHCEASGPRVLSTCRE 3694
QY 671 DCF----FYHEKENQILHYDFSMLSSVSLM-NGPFTSKG 706
DB 3695 GCLCNFGFVLDRDKCVRVECGCKDAQGALIPSGKWTSPG 3735

RESULT 7
TS11_GIALA .STANDARD; PRT; 667 AA.
ID TS11_GIALA .
AC Q03185;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Major surface trophozoite antigen II precursor.
GN TS11.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
OX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Isolate AD-1;
RX MEDLINE=93241215; PubMed=8479449;
RA Ey P.L., Khanna K., Manning P.A., Mayrhofer G.;
RT "A gene encoding a 69-kilodalton major surface protein of Giardia intestinalis trophozoites";
RL Mol. Biochem. Parasitol. 58:247-258(1993).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ANCHORED TO THE PLASMA MEMBRANE.
CC -!- DOMAIN: CONTAINS 27 REPEATS OF A CXXC MOTIF.
CC -!- SIMILARITY: BELONGS TO THE GIARDIA VIRIAN SURFACE PROTEIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).
CC -----
CC EMBL; M95814; AAA02687.1; -
CC PIR; A48579; A48579.
CC HSP; P02468; 1KLO.
CC InterPro; IPR000361; EGF-like.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR005127; Giardia_VSP.
CC Pfam; PF03302; VSP; 1.
CC SMART; SM00181; EGF; 3.
CC SMART; SM00261; FU; 5.
CC Antigen; Repeat; Transmembrane; Signal.
CC SIGNAL 1 17
CC CHAIN 18 667 MAJOR SURFACE TROPHOZOITE ANTIGEN II.
CC DOMAIN 18 633 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 634 662 POTENTIAL.
CC DOMAIN 663 667 CYTOPLASMIC (POTENTIAL).
CC CARBOHYD 591 591 N-LINKED (GLCNAC. .) (POTENTIAL).
CC CARBOHYD 630 630 N-LINKED (GLCNAC. .) (POTENTIAL).
CC SEQUENCE 667 AA; 68475 MW; 1DD9572703232B8D CRC64;

Query Match 3.4%; Score 182; DB 1; Length 667;
Best Local Similarity 19.9%; Pred. No. 8.8e-06;
Matches 146; Conservative 58; Mismatches 236; Indels 292; Gaps 40;

QY 103 CTFSACSGEYLEMKNQVCKGEG-TYISLGSGIKFDEWELPAGFNIAFTMDTVVGPSP 161
DB 141 CSAPKAPGEDNPTKATCTKCAAGFLHTPSEGSLSCSE-ETCPPEGY-----FGHTATAESK 194
```

```
QY 162 SRPDGC--NNSSWIPRGNTVIESNRDDCTVSLIYAVHLKSKGYVFEFYQYVDNNIFFEFTIQ 220
Db 195 KTKCKSTGSGSEAPNVKGI-----GDLCKM-----YNEASGNTL----- 229
QY 221 NDOCEMDTTTDDKWLKLDNGEWSHVMKSGTNILYWRITG-----ILMGSKA 270
Db 230 --TCBKCSAQKPKSLDKTSCNDC-----TGQNCACFCSSGGDCGCGDFILDGQNC 279
QY 271 VRPVLVKNTITIEGVAYTSRCPKPGTKSGFNQCVPCPRNTYSEKGAKE-CIRCKDD 329
Db 280 VK-----SDC-----KTENCRACKT-----NPKAAVEVTECIST 308
QY 330 S0FSSSECTERPPCTKDYFOIHPCDEBQTKQIMYKWIPIKICREDLTDAILPPSGE 389
Db 309 HHLTPTSCQVQY--COALGNYYAGTNADNK-----KACKE--CTVANCKTCD 352
QY 390 KKDCPCPNPGFYNGSS-----SCHPCPPGTFSDGTEKRCPCPA-----GTEPAL 434
Db 353 QGOCOTCNDGFYKNGDACSCHESCKTCSAGTASD-----CTECPGKALKYGNDDGKGTG 408
QY 435 G-----FEYKMNVLPGN-----MKTSCFNVG 456
Db 409 GEGCTGGSGGACKTCGLTIDGASYCSECDTQNEYPQNGICSTTARTVATCKNS--NVA 466
QY 457 NSKCDG-----MNG--WEVAGDHIQSGAGSDNDYLILNLHIPGPKPPTSMGTATGSE 507
Db 467 NGTCSCTNGFLRMNGCGVETTKFKPKSCVCEGANADADTKAPVPGYK-----VE 516
QY 508 LGRITVFETLCSADCVLVFMWDINRKNVSVESGCKTEKQAYTHIPKNTATFTTWAF 567
Db 517 AGKL-----VMGSKG-----DTCSDATTC-----TKCGGYTKI----- 546
QY 568 QRTNQGQDNRREINDMKIYSITATNAVDGVASSCRACALGSEQSGSSCVPCPPGHV--- 624
Db 547 -----EN-----SQTCTKC-----DSSCETCT-----GAATTCVKATGYVYKTA 580
QY 625 IKETNQCCEKCPDPTVLSLTHQVYKEACITPCGPGSKNODHSVCYSDCFYHEKENQILH 684
Db 581 LGEST--CTSCENDS-----NGVIGVKGCLNCAPPS----- 609
QY 685 YDFSNLSSVSGSLMNGPSFTSKTKYFHFFNISLCGHEGKKMALCTNNITDFTVKEIVAGS 744
Db 610 -----SSTGSLV---CYLMKGEN-----TGGSVKNKSLSTGAIAGISVAVVV--- 649
QY 745 DDYTNLVGAFVC 756
Db 650 --VAGLVG-FLC 658
```

RESULT 8.

```
LM2L2_CAEEL STANDARD; PRT; 3672 AA.
AC Q21313;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Laminin-like protein K08C7.3 precursor.
GN K08C7.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderiinae; Caenorhabditis.
ON NCBI_TaxID=6239;
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Berks M.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC -|- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -|- SIMILARITY: CONTAINS 21.5 LAMININ EGF-LIKE DOMAINS.
CC -|- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.
CC -|- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z70286; CAA94293.1;
CC HSSP; P02468; 1K10.
CC WormPep; K08C7.3; CE06136.
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR001886; LamNT.
CC InterPro; IPR000034; Laminin_B.
CC InterPro; IPR002049; Laminin_EGF.
CC InterPro; IPR001791; Laminin_G.
CC Pfam; PF00052; laminin_B; 1.
CC Pfam; PF00053; laminin_EGF; 21.
CC Pfam; PF00054; laminin_G; 5.
CC Pfam; PF00055; laminin_Nterm; 1.
CC PRINTS; PR00011; EGF_LAMININ.
CC ProDom; PD002082; LamNT; 1.
CC ProDom; PD003031; Laminin_B; 1.
CC SMART; SM00180; EGF_Lam; 21.
CC SMART; SM00281; LamB; 1.
CC SMART; SM00282; LamG; 5.
CC SMART; SM00136; LamNT; 1.
CC PROSITE; PS00022; EGF_1; 19.
CC PROSITE; PS01186; EGF_2; 4.
CC PROSITE; PS01248; LAMININ_TYPE_EGF; 21.
CC PROSITE; PS50025; LAM_G_DOMAIN; 5.
CC KW Hypothetical protein; Laminin EGF-like domain; Signal; Repeat.
CC SIGNAL 1 27
CC FT CHAIN 28 3672
CC FT DOMAIN 28 297 LAMININ-LIKE PROTEIN K08C7.3.
CC FT DOMAIN 298 356 LAMININ N-TERMINAL (DOMAIN VI).
CC FT DOMAIN 357 426 LAMININ EGF-LIKE 1.
CC FT DOMAIN 427 471 LAMININ EGF-LIKE 2.
CC FT DOMAIN 472 518 LAMININ EGF-LIKE 3.
CC FT DOMAIN 519 563 LAMININ EGF-LIKE 4.
CC FT DOMAIN 564 609 LAMININ EGF-LIKE 5.
CC FT DOMAIN 610 655 LAMININ EGF-LIKE 6.
CC FT DOMAIN 656 700 LAMININ EGF-LIKE 7.
CC FT DOMAIN 701 755 LAMININ EGF-LIKE 8.
CC FT DOMAIN 756 808 LAMININ EGF-LIKE 9.
CC FT DOMAIN 809 839 LAMININ EGF-LIKE 10.
CC FT DOMAIN 1415 1460 LAMININ EGF-LIKE 11 (INCOMPLETE).
CC FT DOMAIN 1461 1505 LAMININ EGF-LIKE 12.
CC FT DOMAIN 1506 1553 LAMININ EGF-LIKE 13.
CC FT DOMAIN 1554 1604 LAMININ EGF-LIKE 14.
CC FT DOMAIN 1605 1614 LAMININ EGF-LIKE 15.
CC FT DOMAIN 1615 1796 LAMININ EGF-LIKE 16 (N-TERMINAL).
CC FT DOMAIN 1797 1829 LAMININ EGF-LIKE 17.
CC FT DOMAIN 1830 1879 LAMININ EGF-LIKE 18.
CC FT DOMAIN 1880 1936 LAMININ EGF-LIKE 19.
CC FT DOMAIN 1937 1989 LAMININ EGF-LIKE 20.
CC FT DOMAIN 2037 2083 LAMININ EGF-LIKE 21.
CC FT DOMAIN 2084 2131 LAMININ EGF-LIKE 22.
CC FT DOMAIN 2693 2884 LAMININ G-LIKE 1.
CC FT DOMAIN 2896 3056 LAMININ G-LIKE 2.
CC FT DOMAIN 3072 3235 LAMININ G-LIKE 3.
CC FT DOMAIN 3310 3482 LAMININ G-LIKE 4.
CC FT DOMAIN 3488 3669 LAMININ G-LIKE 5.
CC FT DISULFID 298 307 BY SIMILARITY.
CC FT DISULFID 300 320 BY SIMILARITY.
CC FT DISULFID 322 331 BY SIMILARITY.
CC FT DISULFID 334 354 BY SIMILARITY.
CC FT DISULFID 357 366 BY SIMILARITY.
CC FT DISULFID 359 391 BY SIMILARITY.
CC FT DISULFID 394 403 BY SIMILARITY.
CC FT DISULFID 406 424 BY SIMILARITY.
CC FT DISULFID 427 438 BY SIMILARITY.
CC FT DISULFID 429 445 BY SIMILARITY.
```

DISULFID 447 456 BY SIMILARITY.
DISULFID 459 469 BY SIMILARITY.
DISULFID 472 484 BY SIMILARITY.
DISULFID 474 491 BY SIMILARITY.
DISULFID 493 502 BY SIMILARITY.
DISULFID 505 516 BY SIMILARITY.
DISULFID 519 531 BY SIMILARITY.
DISULFID 521 538 BY SIMILARITY.
DISULFID 540 549 BY SIMILARITY.
DISULFID 552 561 BY SIMILARITY.
DISULFID 564 576 BY SIMILARITY.
DISULFID 566 583 BY SIMILARITY.
DISULFID 585 594 BY SIMILARITY.
DISULFID 597 607 BY SIMILARITY.
DISULFID 610 622 BY SIMILARITY.
DISULFID 612 629 BY SIMILARITY.
DISULFID 631 640 BY SIMILARITY.
DISULFID 643 653 BY SIMILARITY.
DISULFID 656 668 BY SIMILARITY.
DISULFID 658 674 BY SIMILARITY.
DISULFID 676 685 BY SIMILARITY.
DISULFID 688 698 BY SIMILARITY.
DISULFID 701 715 BY SIMILARITY.
DISULFID 703 724 BY SIMILARITY.
DISULFID 726 735 BY SIMILARITY.
DISULFID 738 753 BY SIMILARITY.
DISULFID 756 770 BY SIMILARITY.
DISULFID 758 777 BY SIMILARITY.
DISULFID 779 788 BY SIMILARITY.
DISULFID 791 806 BY SIMILARITY.
DISULFID 1415 1427 BY SIMILARITY.
DISULFID 1417 1434 BY SIMILARITY.
DISULFID 1436 1445 BY SIMILARITY.
DISULFID 1448 1458 BY SIMILARITY.
DISULFID 1461 1469 BY SIMILARITY.
DISULFID 1463 1476 BY SIMILARITY.
DISULFID 1478 1487 BY SIMILARITY.
DISULFID 1490 1503 BY SIMILARITY.
DISULFID 1506 1520 BY SIMILARITY.
DISULFID 1508 1527 BY SIMILARITY.
DISULFID 1529 1538 BY SIMILARITY.
DISULFID 1531 1551 BY SIMILARITY.
DISULFID 1554 1566 BY SIMILARITY.
DISULFID 1556 1573 BY SIMILARITY.
DISULFID 1575 1584 BY SIMILARITY.
DISULFID 1587 1602 BY SIMILARITY.
DISULFID 1820 1839 BY SIMILARITY.
DISULFID 1832 1846 BY SIMILARITY.
DISULFID 1849 1858 BY SIMILARITY.
DISULFID 1861 1877 BY SIMILARITY.
DISULFID 1880 1894 BY SIMILARITY.
DISULFID 1882 1905 BY SIMILARITY.
DISULFID 1907 1916 BY SIMILARITY.
DISULFID 1919 1934 BY SIMILARITY.
DISULFID 1937 1951 BY SIMILARITY.
DISULFID 1939 1958 BY SIMILARITY.
DISULFID 1961 1970 BY SIMILARITY.
DISULFID 1973 1987 BY SIMILARITY.
DISULFID 1990 2000 BY SIMILARITY.
DISULFID 1992 2007 BY SIMILARITY.
DISULFID 2009 2018 BY SIMILARITY.
DISULFID 2021 2034 BY SIMILARITY.
DISULFID 2037 2048 BY SIMILARITY.
DISULFID 2039 2055 BY SIMILARITY.
DISULFID 2057 2066 BY SIMILARITY.
DISULFID 2069 2081 BY SIMILARITY.
DISULFID 2084 2096 BY SIMILARITY.
DISULFID 2086 2103 BY SIMILARITY.
DISULFID 2105 2114 BY SIMILARITY.
DISULFID 2117 2129 BY SIMILARITY.
CARBOHYD 121 121 N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD 140 140 N-LINKED (GLCNAC. .) (POTENTIAL).
CARBOHYD 249 249 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 351 351 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 477 477 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 511 511 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 530 530 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 634 634 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 761 761 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1014 1014 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1341 1341 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1705 1705 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1756 1756 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1868 1868 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1944 1944 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1986 1986 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2002 2002 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2159 2159 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2207 2207 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2231 2231 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2235 2235 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2401 2401 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2421 2421 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2487 2487 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2821 2821 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3087 3087 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3242 3242 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3541 3541 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 3672 AA; 404223 MW; 28E262DB5FF14BFA CRC64;
Query Match 3.3%; Score 178.5; DB 1; Length 3672;
Best Local Similarity 22.6%; Pred. No. 0.00014;
Matches 116; Conservative 54; Mismatches 203; Indels 141; Gaps 30;
QY 279 ITIEGVAYTSECPCKP--GTFENKPGSF--NCVCPRTYSEKGAKECIRCKDDSQFSG 334
Db 1413 LSCDCVAQGSSESFOCEYGGQCKCKPGVIGRCERCAPGYN---PPECIKC-----OCNA 1465
QY 335 SSECTERP-PCITTKDYFQIHTPCDEEGKTQIMYKWIPEKI-----CREDLTDAILR--- 384
Db 1466 GOCDERTGOCFCPPHVEGQT-CDRCVSNAGY---DPLIGCKCCHQPGSEGGNLVCD 1521
QY 385 PPSGE-----KKDCPPCNPGFYNNSSSCHPCP---PGT---PSDGTK----- 421
Db 1522 PESGQCLCRESMGGQCDRLAGFY--GPHCYGCSNAGTTEETCDATNAQCKCKENV 1579
QY 422 ---ECRCPAGTEPALGFEYKWNVLPGNMKTSFCNVNKGKCDGMNGWEVAGDHIOGAG 478
Db 1580 YGRCCEACRAGT-----FDLSAENPL-GCVNCFGVTOSCRSSMYPTVIMSDMSSFLT 1633
QY 479 GSDN-----DYLLNLHIPG-----FRPPTSMTGATGSELGRITFVFETL-----C 519
Db 1634 TDDNGMVDNKDDTVIYTSEETSPNSVYFNVPIEKKDYTTSYGLKLTFLKSTVPRGRKSM 1693
QY 520 SADCVLYFMVDINRKSTNV-VESWGG---TKEQAVTHIIFKNATFTFWAFORTNOGOD 575
Db 1694 NAD-----ADVRLTGANMTIEYWASEQTPNEQFT-VKCKLVPENFLTAEKTVTREE 1746
QY 576 NRRFNDMVKI-----YSTATNAVDGV--ASSCRACALGSQSGS 614
Db 1747 LMKVLHSLQNIILKASYFDHPKPTSTLYBFLGLEISEPNGVDVSIKASSVQCQCPAYTGP 1806
QY 615 SCVPCPPGHYIEKTNQCKECPDPTVLSIHOVYKGE---ACIPC---GPGSKNQDHVSC 668
Db 1807 SCQLCASGY-----HRYSGSFLGACVPCPCNGHGHATCDPDGTGIC 1846
QY 669 YSDCFYHEKENOILHYDFSNLSSVGSMLNGPSF 702
Db 1847 -TDC-----EHNNTNGDHCEFCNEGHYGNATNGSPY 1875
RESULT 9
LMGL_MOUSE
ID LMGL_MOUSE STANDARD; PRT; 1607 AA.
AC P02468;
DT 21-JUL-1986 (Rel. 01, Created)

DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Laminin gamma-1 chain precursor (Laminin B2 chain).
 GN LAMC1 OR LAMC-1 OR LAMB-2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88059118; PubMed=3680290;
 RA Sasaki M., Yamada Y.;
 RT "The laminin B2 chain has a multidomain structure homologous to the
 RT B1 chain.";
 RL J. Biol. Chem. 262:17111-17117(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89000737; PubMed=3167041;
 RA Durkin M.E., Bartos B.B., Liu S.-H., Phillips S.L., Chung A.E.;
 RT "Primary structure of the mouse laminin B2 chain and comparison with
 RT laminin B1.";
 RL Biochemistry 27:5198-5204(1988).
 RN [3]
 RP SEQUENCE OF 1-239 FROM N.A.
 RX MEDLINE=88228071; PubMed=2836421;
 RA Ogawa K., Burdello P.D., Sasaki M., Yamada Y.;
 RT "The laminin B2 chain promoter contains unique repeat sequences and
 RT is active in transient transfection.";
 RL J. Biol. Chem. 263:8384-8389(1988).
 RN [4]
 RP SEQUENCE OF 1391-1607 FROM N.A.
 RX MEDLINE=85051302; PubMed=6209134;
 RA Barlow D.P., Green N.M., Kurkinen M., Hogan B.L.M.;
 RT "Sequencing of laminin B chain cDNAs reveals C-terminal regions of
 RT coiled-coil alpha-helix.";
 RL EMBO J. 3:2355-2362(1984).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 771-932.
 RX MEDLINE=96196434; PubMed=8648630;
 RA Stetefeld J., Mayer U., Timpl R., Huber R.;
 RT "Crystal structure of three consecutive laminin-type epidermal growth
 RT factor-like (LE) modules of laminin gammal chain harboring the
 RT nidogen binding site.";
 RL J. Mol. Biol. 257:644-657(1996).
 RN [6]
 RP STRUCTURE BY NMR OF 824-881.
 RX MEDLINE=96196435; PubMed=8648631;
 RA Baumgartner R., Czisch M., Mayer U., Poeschl E., Huber R.,
 RA Timpl R., Holak T.A.;
 RT "Structure of the nidogen binding LE module of the laminin gammal
 RT chain in solution.";
 RL J. Mol. Biol. 257:658-668(1996).
 CC -!- FUNCTION: Binding to cells via a high affinity receptor, laminin
 CC is thought to mediate the attachment, migration, and organization
 CC of cells into tissues during embryonic development by interacting
 CC with other extracellular matrix components.
 CC -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
 CC different polypeptide chains (alpha, beta, gamma), which are bound
 CC to each other by disulfide bonds into a cross-shaped molecule
 CC comprising one long and three short arms with globules at each
 CC end.
 CC THE GAMMA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ),
 CC LAMININ-2 (MEROSIN), LAMININ-3 (S-LAMININ), LAMININ-4 (S-MEROSIN),
 CC LAMININ-6 (K-LAMININ) AND LAMININ-7 (KS-LAMININ).
 CC -!- SUBCELLULAR LOCATION: Extracellular.
 CC -!- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
 CC COMPONENT).
 CC -!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
 CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
 CC -!- DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.
 CC -!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
 CC -!- SIMILARITY: CONTAINS 11 LAMININ EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X05211; CAA28838.1; -
 CC EMBL: J03484; AAA39405.1; -
 CC EMBL: J02930; AAA39408.1; -
 CC EMBL: J03749; AAA39409.1; -
 CC PIR: A28469; MMSB2.
 CC PDB: 1KLO; 20-AUG-97.
 CC PDB: 1TLE; 12-FEB-97.
 CC MGI: 99914; Lamc1.
 CC InterPro: IPR004089; Chmtaxis_transd.
 CC InterPro: IPR000561; EGF-like.
 CC InterPro: IPR001886; LamNT.
 CC InterPro: IPR000034; Laminin-B.
 CC InterPro: IPR002049; Laminin_EGF.
 CC Pfam: PF00052; laminin-B; 1.
 CC Pfam: PF00053; laminin_EGF; 10.
 CC PRINTS: PR00011; EGFLAMININ.
 CC ProDom: PD002082; LamNT; 1.
 CC ProDom: PD003031; Laminin-B; 1.
 CC SMART: SM00180; EGF_Lam; 9.
 CC SMART: SM00001; EGF_Like; 1.
 CC SMART: SM00281; LamB; 1.
 CC SMART: SM00136; LamNT; 1.
 CC PROSITE: PS00022; EGF_1; 8.
 CC PROSITE: PS01186; EGF_2; 2.
 CC PROSITE: PS01248; LAMININ_TYPE_EGF; 10.
 CC Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
 CC Laminin EGF-like domain; Cell adhesion; Repeat; Signal; 3D-structure.
 CC SIGNAL 1 33
 CC CHAIN 34 1607 LAMININ GAMMA-1 CHAIN.
 CC DOMAIN 34 283 LAMININ N-TERMINAL (DOMAIN VI).
 CC DOMAIN 284 339 LAMININ EGF-LIKE 1.
 CC DOMAIN 340 395 LAMININ EGF-LIKE 2.
 CC DOMAIN 396 442 LAMININ EGF-LIKE 3.
 CC DOMAIN 443 492 LAMININ EGF-LIKE 4.
 CC DOMAIN 493 502 LAMININ EGF-LIKE 5 (N-TERMINAL).
 CC DOMAIN 503 687 LAMININ DOMAIN IV.
 CC DOMAIN 688 721 LAMININ EGF-LIKE 5 (C-TERMINAL).
 CC DOMAIN 722 770 LAMININ EGF-LIKE 6.
 CC DOMAIN 771 825 LAMININ EGF-LIKE 7.
 CC DOMAIN 826 881 LAMININ EGF-LIKE 8 (NTDOGEN-BINDING).
 CC DOMAIN 882 932 LAMININ EGF-LIKE 9.
 CC DOMAIN 933 980 LAMININ EGF-LIKE 10.
 CC DOMAIN 981 1028 LAMININ EGF-LIKE 11.
 CC DOMAIN 1029 1607 DOMAIN II AND I.
 CC DOMAIN 1034 1594 COILED COIL (POTENTIAL).
 CC DISULFID 340 349 BY SIMILARITY.
 CC DISULFID 342 365 BY SIMILARITY.
 CC DISULFID 368 377 BY SIMILARITY.
 CC DISULFID 380 393 BY SIMILARITY.
 CC DISULFID 396 408 BY SIMILARITY.
 CC DISULFID 398 414 BY SIMILARITY.
 CC DISULFID 416 425 BY SIMILARITY.
 CC DISULFID 428 440 BY SIMILARITY.
 CC DISULFID 443 454 BY SIMILARITY.
 CC DISULFID 445 461 BY SIMILARITY.
 CC DISULFID 463 472 BY SIMILARITY.
 CC DISULFID 475 490 BY SIMILARITY.
 CC DISULFID 722 731 BY SIMILARITY.
 CC DISULFID 724 738 BY SIMILARITY.
 CC DISULFID 740 749 BY SIMILARITY.
 CC DISULFID 752 768 BY SIMILARITY.
 CC DISULFID 771 779 BY SIMILARITY.
 CC DISULFID 773 790 BY SIMILARITY.

RT "Developmental expression of fibrillin genes suggests heterogeneity
 of extracellular microfibrils";
 J. Cell Biol. 129:1165-1176(1995).
 [2]
 RP SEQUENCE OF 210-317 FROM N.A.
 RX MEDLINE=94140368; PubMed=8307578;
 RA Li X., Pereira L., Zhang H., Sanguinetti C., Ramirez F., Bonadio J.,
 Francke U.,
 RT "Fibrillin genes map to regions of conserved mouse/human synteny on
 mouse chromosomes 2 and 18";
 RL Genomics 18:667-672(1993).
 CC -!- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS
 THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE
 CC LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.
 CC -!- SIMILARITY: CONTAINS 47 EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L39790; AAA74908.1; -;
 DR EMBL: S69359; AAC60685.1; -;
 DR HSSP: P35355; 1EMN.
 DR MGD: MGI:95490; Fbn2.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF Ca.
 DR InterPro: IPR001438; EGF II.
 DR InterPro: IPR002212; Fibril-assoc.
 DR Pfam: PF00008; EGF; 46.
 DR Pfam: PF00683; TB; 9.
 DR PRINTS: PRO0010; EGFBLD.
 DR SMART: SM00179; EGF_CA; 43.
 DR SMART: SM00001; EGF-like; 3.
 DR PROSITE: PS00010; ASX_HYDROXYL; 43.
 DR PROSITE: PS00022; EGF_1; 2.
 DR PROSITE: PS01186; EGF_2; 36.
 DR PROSITE: PS01187; EGF_CA; 43.
 DR Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;
 KW Repeat; Signal; Multigene family.
 FT SIGNAL 1 28 POTENTIAL.
 FT CHAIN 29 2907 FIBRILLIN 2.
 FT DOMAIN 111 142 EGF-LIKE 1.
 FT DOMAIN 145 176 EGF-LIKE 2.
 FT DOMAIN 176 208 EGF-LIKE 3.
 FT DOMAIN 276 317 EGF-LIKE 4, CALCIUM-BINDING.
 FT DOMAIN 318 359 EGF-LIKE 5, CALCIUM-BINDING.
 FT REPEAT 360 426 TGFBP 1.
 FT DOMAIN 487 527 EGF-LIKE 6.
 FT DOMAIN 528 567 EGF-LIKE 7, CALCIUM-BINDING.
 FT DOMAIN 568 609 EGF-LIKE 8, CALCIUM-BINDING.
 FT DOMAIN 610 650 EGF-LIKE 9, CALCIUM-BINDING.
 FT DOMAIN 651 691 EGF-LIKE 10, CALCIUM-BINDING.
 FT REPEAT 692 760 TGFBP 2.
 FT DOMAIN 761 802 EGF-LIKE 11, CALCIUM-BINDING.
 FT DOMAIN 803 844 EGF-LIKE 12, CALCIUM-BINDING.
 FT DOMAIN 845 883 EGF-LIKE 13, CALCIUM-BINDING.
 FT DOMAIN 948 989 EGF-LIKE 14, CALCIUM-BINDING.
 FT REPEAT 990 1065 TGFBP 3.
 FT DOMAIN 1066 1107 EGF-LIKE 15, CALCIUM-BINDING.
 FT DOMAIN 1108 1150 EGF-LIKE 16, CALCIUM-BINDING.
 FT DOMAIN 1151 1192 EGF-LIKE 17, CALCIUM-BINDING.
 FT DOMAIN 1193 1234 EGF-LIKE 18, CALCIUM-BINDING.
 FT DOMAIN 1235 1275 EGF-LIKE 19, CALCIUM-BINDING.
 FT DOMAIN 1276 1317 EGF-LIKE 20, CALCIUM-BINDING.
 FT DOMAIN 1318 1359 EGF-LIKE 21, CALCIUM-BINDING.
 FT DOMAIN 1360 1400 EGF-LIKE 22, CALCIUM-BINDING.
 FT DOMAIN 1401 1441 EGF-LIKE 23, CALCIUM-BINDING.

FT DOMAIN 1442 1483 EGF-LIKE 24, CALCIUM-BINDING.
 FT DOMAIN 1484 1524 EGF-LIKE 25, CALCIUM-BINDING.
 FT DOMAIN 1525 1565 EGF-LIKE 26, CALCIUM-BINDING.
 FT REPEAT 1566 1642 TGFBP 4.
 FT DOMAIN 1684 1726 EGF-LIKE 27, CALCIUM-BINDING.
 FT DOMAIN 1727 1800 TGFBP 5.
 FT DOMAIN 1801 1842 EGF-LIKE 29, CALCIUM-BINDING.
 FT DOMAIN 1843 1884 EGF-LIKE 30, CALCIUM-BINDING.
 FT DOMAIN 1885 1926 EGF-LIKE 31, CALCIUM-BINDING.
 FT DOMAIN 1927 1965 EGF-LIKE 32, CALCIUM-BINDING.
 FT DOMAIN 1966 2008 EGF-LIKE 33, CALCIUM-BINDING.
 FT DOMAIN 2009 2048 EGF-LIKE 34, CALCIUM-BINDING.
 FT DOMAIN 2049 2090 EGF-LIKE 35, CALCIUM-BINDING.
 FT REPEAT 2091 2163 TGFBP 6.
 FT DOMAIN 2164 2205 EGF-LIKE 36, CALCIUM-BINDING.
 FT DOMAIN 2206 2245 EGF-LIKE 37, CALCIUM-BINDING.
 FT DOMAIN 2246 2286 EGF-LIKE 38, CALCIUM-BINDING.
 FT DOMAIN 2287 2330 EGF-LIKE 39, CALCIUM-BINDING.
 FT DOMAIN 2331 2372 EGF-LIKE 40, CALCIUM-BINDING.
 FT REPEAT 2373 2441 TGFBP 7.
 FT DOMAIN 2442 2483 EGF-LIKE 41, CALCIUM-BINDING.
 FT DOMAIN 2484 2524 EGF-LIKE 42, CALCIUM-BINDING.
 FT DOMAIN 2525 2564 EGF-LIKE 43, CALCIUM-BINDING.
 FT DOMAIN 2565 2606 EGF-LIKE 44, CALCIUM-BINDING.
 FT DOMAIN 2607 2646 EGF-LIKE 45, CALCIUM-BINDING.
 FT DOMAIN 2647 2687 EGF-LIKE 46, CALCIUM-BINDING.
 FT DOMAIN 2688 2727 EGF-LIKE 47, CALCIUM-BINDING.
 FT DISULFID 115 124 BY SIMILARITY.
 FT DISULFID 119 130 BY SIMILARITY.
 FT DISULFID 132 141 BY SIMILARITY.
 FT DISULFID 149 159 BY SIMILARITY.
 FT DISULFID 153 164 BY SIMILARITY.
 FT DISULFID 166 175 BY SIMILARITY.
 FT DISULFID 180 190 BY SIMILARITY.
 FT DISULFID 184 196 BY SIMILARITY.
 FT DISULFID 198 207 BY SIMILARITY.
 FT DISULFID 280 292 BY SIMILARITY.
 FT DISULFID 287 301 BY SIMILARITY.
 FT DISULFID 303 316 BY SIMILARITY.
 FT DISULFID 322 334 BY SIMILARITY.
 FT DISULFID 329 343 BY SIMILARITY.
 FT DISULFID 345 358 BY SIMILARITY.
 FT DISULFID 491 503 BY SIMILARITY.
 FT DISULFID 498 512 BY SIMILARITY.
 FT DISULFID 514 526 BY SIMILARITY.
 FT DISULFID 532 542 BY SIMILARITY.
 FT DISULFID 537 551 BY SIMILARITY.
 FT DISULFID 553 566 BY SIMILARITY.
 FT DISULFID 572 584 BY SIMILARITY.
 FT DISULFID 579 593 BY SIMILARITY.
 FT DISULFID 595 608 BY SIMILARITY.
 FT DISULFID 614 625 BY SIMILARITY.
 FT DISULFID 620 634 BY SIMILARITY.
 FT DISULFID 636 649 BY SIMILARITY.
 FT DISULFID 655 666 BY SIMILARITY.
 FT DISULFID 661 675 BY SIMILARITY.
 FT DISULFID 677 690 BY SIMILARITY.
 FT DISULFID 765 777 BY SIMILARITY.
 FT DISULFID 772 786 BY SIMILARITY.
 FT DISULFID 788 801 BY SIMILARITY.
 FT DISULFID 807 819 BY SIMILARITY.
 FT DISULFID 814 828 BY SIMILARITY.
 FT DISULFID 830 843 BY SIMILARITY.
 FT DISULFID 849 859 BY SIMILARITY.
 FT DISULFID 854 868 BY SIMILARITY.
 FT DISULFID 870 883 BY SIMILARITY.
 FT DISULFID 952 964 BY SIMILARITY.
 FT DISULFID 959 973 BY SIMILARITY.
 FT DISULFID 975 988 BY SIMILARITY.
 FT DISULFID 1070 1082 BY SIMILARITY.
 FT DISULFID 1077 1091 BY SIMILARITY.
 FT DISULFID 1093 1106 BY SIMILARITY.

FT	DISULFID	1112	1124	BY SIMILARITY.
FT	DISULFID	1119	1133	BY SIMILARITY.
FT	DISULFID	1135	1149	BY SIMILARITY.
FT	DISULFID	1155	1167	BY SIMILARITY.
FT	DISULFID	1162	1176	BY SIMILARITY.
FT	DISULFID	1178	1191	BY SIMILARITY.
FT	DISULFID	1197	1209	BY SIMILARITY.
FT	DISULFID	1204	1218	BY SIMILARITY.
FT	DISULFID	1220	1233	BY SIMILARITY.
FT	DISULFID	1239	1250	BY SIMILARITY.
FT	DISULFID	1246	1259	BY SIMILARITY.
FT	DISULFID	1261	1274	BY SIMILARITY.
FT	DISULFID	1280	1292	BY SIMILARITY.
FT	DISULFID	1287	1301	BY SIMILARITY.
FT	DISULFID	1303	1316	BY SIMILARITY.
FT	DISULFID	1322	1334	BY SIMILARITY.
FT	DISULFID	1329	1343	BY SIMILARITY.
FT	DISULFID	1345	1358	BY SIMILARITY.
FT	DISULFID	1364	1377	BY SIMILARITY.
FT	DISULFID	1371	1386	BY SIMILARITY.
FT	DISULFID	1388	1399	BY SIMILARITY.
FT	DISULFID	1405	1418	BY SIMILARITY.
FT	DISULFID	1412	1427	BY SIMILARITY.
FT	DISULFID	1429	1440	BY SIMILARITY.
FT	DISULFID	1446	1458	BY SIMILARITY.
FT	DISULFID	1453	1467	BY SIMILARITY.
FT	DISULFID	1469	1482	BY SIMILARITY.
FT	DISULFID	1488	1499	BY SIMILARITY.
FT	DISULFID	1494	1508	BY SIMILARITY.
FT	DISULFID	1510	1523	BY SIMILARITY.
FT	DISULFID	1529	1540	BY SIMILARITY.
FT	DISULFID	1535	1549	BY SIMILARITY.
FT	DISULFID	1551	1564	BY SIMILARITY.
FT	DISULFID	1647	1659	BY SIMILARITY.
FT	DISULFID	1654	1668	BY SIMILARITY.
FT	DISULFID	1670	1683	BY SIMILARITY.
FT	DISULFID	1689	1701	BY SIMILARITY.
FT	DISULFID	1696	1710	BY SIMILARITY.
FT	DISULFID	1712	1725	BY SIMILARITY.

Dn	335	--SNTVGSYFC-LCPRGFTVTSDSR-CIDQAGTCFSGLVNRCQAQLPGRMAKAQC	393
Qy	452	-----CFWGN--SKC-----DGMNGWEVAGDHQ--SGAGSSNDYLILN-488	
Dn	390	CEPGRCMSIGTIPACPVRSSEYYRRCLDLGLPMGGITPGSVSRPCTGSGNGYGPGGT	449
Qy	489	--LHIP---GFKPPTSMTGATSELGRI---TFVFET-----LC-----519	
Dn	450	GFLPIPDGNSPGVGAGVAGGOGPIITGLTILNOTIIDICKHAMLNLUNGRCIPTVSS	509
Qy	520	-----SADCVLYFMVDINRKSNNVVESMGTTKEKAY---THIIFKNATFTF	563
Dn	510	YRCRCNMGYKODANGDC-----IDVDECTSNPCSGNGDCVNTPGSYCKCHA-----555	
Qy	564	TWAQTNOGQ-----DNRRFINDMWKIYSI-----TATNAVD-GV	598
Dn	556	--GFRPTPKOACIDIDECIQNGVLCKNGRCVNSDGSCFOCICNAGFELTDDGKNCDVIDE	613
Qy	599	ASSCRACALG---SEQSGSSVCPCPPGHYTEKETNOC---KEC-PPDTYLSIHGV-YGK	649
Dn	614	CTTNMCLNGMCINEDSGSKVC-KPGFILAPNGRYCTDVDECOTPGICNHGHCINNIGS	672
Qy	650	EACIPCGPGSKNNODHSVCY-----SDCFYHEKENQILHYDFSJNLSSVGSMLNGPSF--	702
Dn	673	PRC-DCPPLGAVGDGRVCDTHMRSTCYGEIKKGVCVRPFPAVTKSECCAN-PDYGF	730
Qy	703	-----TSKGTKYPH-----FFNISLCGHEGKMAL-----CNNTITDFTVKIVAGSD	746
Dn	731	GPCPCPAKNASAEFHGLCSGIGITVDGRDINECALDPDICANGICE-----778	
Qy	747	YTNLGVAFVCOSTIIPSESKFRAALSQSISLIADTFIGVTVETTLKNIKED-----800	
Dn	779	--NLRGHYRNC-----NSGYEPDASGRNCIDIECL---VNRLCDNGLCRNTPGSGYS	827
Qy	801	-----MFVPVPTSQIPDVHFYFKSSTATTCINGRST----AVKRCRNPTK--SGAGVI	847
Dn	828	CTCPPGGVYLPTEETCEDV-----NECESNPCVNGACRNNLGSHFCHBCSPGSLSTGLI	882
Qy	848	SVPKSCPATC-----DGCITFYF-----LWESAECAPLICTEHDFHEIEG	886
Dn	883	CIDSU--KGTCWLNIQNRCEVNINGATLKSECCATLGGAWGS--PCERC-----ELDA	932
Qy	887	ACKRGFOETLYVMNEPKWCIGISLPEKKLATCETVD	923
Dn	932	ACPRGAR-----IKGV-----TCEDVN	950
RESULT 11			
LMG3_MOUSE			
ID	LMG3_MOUSE	STANDARD;	PRT; 1581 AA.
AC	O9RB6; O9WTW6;		
DT	15-JUN-2002 (Rel. 41, Created)		
DT	15-JUN-2002 (Rel. 41, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Laminin gamma-3 chain precursor (Laminin 12 gamma 3).		
LN	LAMC3.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OX	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
ON	NCBI_TaxId=10090;		
RN	[1]		
RN	SEQUENCE FROM N.A.		
RA	Albus A.M., Burgeson B., Champlaud M.-F., Koch M., Olson P.;		
RT	"Mouse laminin 12 gamma 3 chain."		
.RL	Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RN	SEQUENCE OF 1-1526 FROM N.A.		
RP	MEDLINE=99253969; PubMed=10318827;		
RX	Iivanainen A., Morita T.; Tryggvason K.;		
RA	"Molecular cloning and tissue-specific expression of a novel murine		
RT	laminin gamma3 chain,"		
RL	J. Biol. Chem. 274:14107-14111(1999).		


```
QY 615 SCVPCPPGHY---IEKETNCKECP-----PDTYLSIHQVYVYKACIPCGPGSKN 661
DB 744 SCERCMPGFYNAFSGRADDCPCPGOSACATIPES-----GDVCTHCPPGQRG 795
QY 662 NODHSVCYSDCFYHEKENQILHYDFSLSVGLNMGSPFTSKGTYKHYHFFNISLCGHE 721
DB 796 RCSES-C-EDGFF-----GDLGLSGAP-----O 817
QY 722 GKMACTNNITDFTVKEIVAGSDDDY-----NLVGAFCQSTIIPSESKGFRALIS 773
DB 818 PCRRCCSGNVD---LNAVGNCDPHSGHCLRLYNTTGAH-CEH-----857
QY 774 SOSIILADFIQVVTTLKNNIKEDMPVPVTSQIPDVHFFYKSTATTSCINGRSTAV 833
DB 858 -----CREGFYGSAVATR-----PV-----DKAPCSC-DLRGSVS 887
QY 834 KMRCNPTKSGAGVISVP-----SKPAGTCGCTFYFLWESAEACPLCTEHDHFEIEG 886
DB 888 EKTCTNPTVTCQ--VCLPYVSGRDCSRCSPG-----FYDL-QSGRGCSQCKCHPLGSLN 938
QY 887 AC 888
DB 939 KC 940

RESULT 12
FBN1_MOUSE
ID FBN1_MOUSE STANDARD; PRT; 2871 AA.
AC Q61554; Q60826;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Fibrillin 1 precursor.
GN FBN1 OR FBN-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=95130561; PubMed=7829516;
RA Yin W., Gemiller J., Sanguinetti C., Smiley E., Pangilinan T.,
RA Pereira L., Ramirez F., Bonadio J.;
RT "Primary structure and developmental expression of Fbn-1, the mouse
RT fibrillin gene.";
RL J. Biol. Chem. 270:1798-1806(1995).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=CD-1; TISSUE=Kidney;
RA Ota K., Kumar A., Wada J., Liu Z., Kanwar Y.S.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS
CC THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE
CC LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.
CC -!- PTM: FORMS INTERMOLECULAR DISULFIDE BONDS EITHER WITH OTHER
CC FIBRILLIN-1 MOLECULES OR WITH OTHER COMPONENTS OF THE
CC MICROFIBRILS (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 47 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L29454; AAA56840.1; -
CC EMBL; U22493; AAA64217.1; -
CC HSSP; P35555; IAPJ.
CC MGD; MGI:95489; Fbn1.
DR InterPro; IPR000152; Asx_hydroxyl.
```

```
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR002212; Fibril-assoc.
DR Pfam; PF00008; EGF; 46.
DR Pfam; PF00683; TB; 9.
DR SMART; SM00179; EGF_CA; 42.
DR SMART; SM00001; EGF_like; 4.
DR PROSITE; PS00010; ASX_HYDROXYL; 43.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 38.
DR PROSITE; PS01187; EGF_CA; 43.
KW Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;
KW Repeat; Signal; Multigene family.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 2871 FIBRILLIN 1.
FT DOMAIN 91 112 EGF-LIKE 1.
FT DOMAIN 115 146 EGF-LIKE 2.
FT DOMAIN 147 178 EGF-LIKE 3.
FT DOMAIN 246 287 EGF-LIKE 4.
FT DOMAIN 288 329 EGF-LIKE 5, CALCIUM-BINDING.
FT DOMAIN 330 401 TGFBP 1.
FT DOMAIN 402 446 PRO-RICH.
FT DOMAIN 449 489 EGF-LIKE 6.
FT DOMAIN 490 529 EGF-LIKE 7, CALCIUM-BINDING.
FT DOMAIN 530 571 EGF-LIKE 8, CALCIUM-BINDING.
FT DOMAIN 572 612 EGF-LIKE 9, CALCIUM-BINDING.
FT DOMAIN 613 653 EGF-LIKE 10, CALCIUM-BINDING.
FT DOMAIN 656 721 TGFBP 2.
FT DOMAIN 723 764 EGF-LIKE 11, CALCIUM-BINDING.
FT DOMAIN 765 806 EGF-LIKE 12, CALCIUM-BINDING.
FT DOMAIN 807 846 EGF-LIKE 13, CALCIUM-BINDING.
FT DOMAIN 910 951 EGF-LIKE 14, CALCIUM-BINDING.
FT DOMAIN 952 1018 TGFBP 3.
FT DOMAIN 1028 1069 EGF-LIKE 15, CALCIUM-BINDING.
FT DOMAIN 1070 1112 EGF-LIKE 16, CALCIUM-BINDING.
FT DOMAIN 1113 1154 EGF-LIKE 17, CALCIUM-BINDING.
FT DOMAIN 1155 1196 EGF-LIKE 18, CALCIUM-BINDING.
FT DOMAIN 1197 1237 EGF-LIKE 19, CALCIUM-BINDING.
FT DOMAIN 1238 1279 EGF-LIKE 20, CALCIUM-BINDING.
FT DOMAIN 1280 1321 EGF-LIKE 21, CALCIUM-BINDING.
FT DOMAIN 1322 1362 EGF-LIKE 22, CALCIUM-BINDING.
FT DOMAIN 1363 1403 EGF-LIKE 23, CALCIUM-BINDING.
FT DOMAIN 1404 1445 EGF-LIKE 24, CALCIUM-BINDING.
FT DOMAIN 1446 1486 EGF-LIKE 25, CALCIUM-BINDING.
FT DOMAIN 1487 1527 EGF-LIKE 26, CALCIUM-BINDING.
FT DOMAIN 1528 1559 TGFBP 4.
FT DOMAIN 1606 1647 EGF-LIKE 27, CALCIUM-BINDING.
FT DOMAIN 1648 1688 EGF-LIKE 28, CALCIUM-BINDING.
FT DOMAIN 1689 1758 TGFBP 5.
FT DOMAIN 1766 1807 EGF-LIKE 29, CALCIUM-BINDING.
FT DOMAIN 1808 1848 EGF-LIKE 30, CALCIUM-BINDING.
FT DOMAIN 1849 1890 EGF-LIKE 31, CALCIUM-BINDING.
FT DOMAIN 1891 1929 EGF-LIKE 32, CALCIUM-BINDING.
FT DOMAIN 1930 1972 EGF-LIKE 33, CALCIUM-BINDING.
FT DOMAIN 1973 2012 EGF-LIKE 34, CALCIUM-BINDING.
FT DOMAIN 2013 2054 EGF-LIKE 35, CALCIUM-BINDING.
FT DOMAIN 2055 2121 TGFBP 6.
FT DOMAIN 2122 2165 EGF-LIKE 36, CALCIUM-BINDING.
FT DOMAIN 2166 2205 EGF-LIKE 37, CALCIUM-BINDING.
FT DOMAIN 2206 2246 EGF-LIKE 38, CALCIUM-BINDING.
FT DOMAIN 2247 2290 EGF-LIKE 39, CALCIUM-BINDING.
FT DOMAIN 2291 2332 EGF-LIKE 40, CALCIUM-BINDING.
FT DOMAIN 2333 2400 TGFBP 7.
FT DOMAIN 2402 2443 EGF-LIKE 41, CALCIUM-BINDING.
FT DOMAIN 2444 2484 EGF-LIKE 42, CALCIUM-BINDING.
FT DOMAIN 2485 2523 EGF-LIKE 43, CALCIUM-BINDING.
FT DOMAIN 2524 2566 EGF-LIKE 44, CALCIUM-BINDING.
FT DOMAIN 2567 2606 EGF-LIKE 45, CALCIUM-BINDING.
FT DOMAIN 2607 2647 EGF-LIKE 46, CALCIUM-BINDING.
FT DOMAIN 2648 2687 EGF-LIKE 47, CALCIUM-BINDING.
FT DISULFID 85 94 BY SIMILARITY.
FT DISULFID 99 100 BY SIMILARITY.
FT DISULFID 102 111 BY SIMILARITY.
```


QY 735 FTVKEI----VAGSDYTNLGAFCVQSTIIPSEKGFRAALSSQSIILADFFIGTWTET 790
 DB 1728 ANNKCEQCPISPTDEFAFLCG-----SORPGF-----VIDIYGLPVD- 1766
 QY 791 TLKNINIKEDMFPVTSQIP-----DVHFFYKSSSTAT-----TSCIN 827
 DB 1767 -----IDE-----CREIPGVCNGVCINMVGSPFCPCVGGFFYNDKLLVGCIDECON 1814
 QY 828 GR-----STAVKMRNPTKSGAGVIS-----VPSKCPAGTCDGC--TFY 864
 DB 1815 GPVLRLNAECINTAGSYRCD-CKPYRLTSTGQCNDRNECQEPINCSHGQCIDTVGSGFY 1873
 QY 865 FL-----WSAEACPLCTEHDHFIEE-GACKRG 891
 DB 1874 CLCHTGFKTNEQDTMCL--DINECERDCCGNG 1903

RESULT 13
 LMB1_MOUSE
 ID LMB1_MOUSE STANDARD; PRT; 1786 AA.
 AC P02469;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Laminin beta-1 chain precursor (Laminin B1 chain).
 GN LMB1-1 OR LAMB-1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP MEDLINE=87147212; PubMed=3493487;
 RA Sasaki M., Kato S., Kohno K., Martin G.R., Yamada Y.;
 RT "Sequence of the cDNA encoding the laminin B1 chain reveals a
 multimeric protein containing cysteine-rich repeats.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:935-939(1987).
 RN [2]
 RP SEQUENCE OF 1292-1786 FROM N.A.
 RA MEDLINE=85051302; PubMed=6209134;
 RA Barlow D.P., Green N.M., Kurkinen M., Hogan B.L.M.;
 RT "Sequencing of laminin B chain cDNAs reveals C-terminal regions of
 coiled-coil alpha-helix.";
 RL EMBO J. 3:2355-2362(1984).
 RN [3]
 RP SEQUENCE OF 165-172; 539-547 AND 712-719.
 RC STRAIN=BALB/c; TISSUE=Endothelial cells;
 RX MEDLINE=97363207; PubMed=9219532;
 RA Frieser M., Noeckel H., Pausch F., Roeder C., Hahn A., Deutzmann R.,
 RA Sorokin L.M.;
 RT "Cloning of the mouse laminin alpha 4 cDNA. Expression in a subset of
 endothelium.";
 RL Eur. J. Biochem. 246:727-735(1997).
 CC -!- FUNCTION: Binding to cells via a high affinity receptor, laminin
 is thought to mediate the attachment, migration, and organization
 of cells into tissues during embryonic development by interacting
 with other extracellular matrix components.
 CC -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
 different polypeptide chains (alpha, beta, gamma), which are bound
 to each other by disulfide bonds into a cross-shaped molecule
 comprising one long and three short arms with globules at each
 end.
 CC THE BETA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (BHS LAMININ), LAMININ-
 2 (MEROSIN), AND LAMININ-6 (K-LAMININ).
 CC -!- SUBCELLULAR LOCATION: Extracellular.
 CC -!- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
 COMPONENT).
 CC -!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
 CC -!- SIMILARITY: CONTAINS 12.5 LAMININ EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M15525; AAA39407.1; ALT_INIT.
 DR EMBL; X05212; CAA28839.1;
 DR PIR; A26413; MNMSB1.
 DR HSP; P02468; IKLO.
 DR MGD; MGI:96743; Lmb1-1.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001886; LamNT.
 DR InterPro; IPR002049; Laminin_EGF.
 DR Pfam; PF00053; laminin_EGF; 13.
 DR Pfam; PF00055; laminin_Nterm; 1.
 DR PRINTS; PR00011; EGF_LAMININ.
 DR ProDom; PD02082; LamNT; 1.
 DR SMART; SM00180; EGF_Lam; 11.
 DR SMART; SM00136; LamNT; 1.
 DR PROSITE; PS00022; EGF_1; 9.
 DR PROSITE; PS01186; EGF_2; 2.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
 KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
 KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
 FT SIGNAL 1 21
 FT CHAIN 22 1786 LAMININ BETA-1 CHAIN.
 FT DOMAIN 22 270 LAMININ N-TERMINAL (DOMAIN VI).
 FT DOMAIN 271 334 LAMININ EGF-LIKE 1.
 FT DOMAIN 335 397 LAMININ EGF-LIKE 2.
 FT DOMAIN 398 457 LAMININ EGF-LIKE 3.
 FT DOMAIN 458 509 LAMININ EGF-LIKE 4.
 FT DOMAIN 510 540 LAMININ EGF-LIKE 5 (INCOMPLETE).
 FT DOMAIN 541 772 LAMININ DOMAIN IV.
 FT DOMAIN 773 820 LAMININ EGF-LIKE 6.
 FT DOMAIN 821 866 LAMININ EGF-LIKE 7.
 FT DOMAIN 867 916 LAMININ EGF-LIKE 8.
 FT DOMAIN 917 975 LAMININ EGF-LIKE 9.
 FT DOMAIN 976 1027 LAMININ EGF-LIKE 10.
 FT DOMAIN 1028 1083 LAMININ EGF-LIKE 11.
 FT DOMAIN 1084 1131 LAMININ EGF-LIKE 12.
 FT DOMAIN 1132 1178 LAMININ EGF-LIKE 13.
 FT DOMAIN 1179 1397 DOMAIN II.
 FT DOMAIN 1398 1430 DOMAIN ALPHA.
 FT DOMAIN 1431 1786 DOMAIN I.
 FT DOMAIN 1216 1315 COILED COIL (POTENTIAL).
 FT DOMAIN 1368 1388 COILED COIL (POTENTIAL).
 FT DOMAIN 1448 1778 COILED COIL (POTENTIAL).
 FT DISULFID 271 280 BY SIMILARITY.
 FT DISULFID 273 298 BY SIMILARITY.
 FT DISULFID 300 309 BY SIMILARITY.
 FT DISULFID 312 332 BY SIMILARITY.
 FT DISULFID 335 344 BY SIMILARITY.
 FT DISULFID 337 362 BY SIMILARITY.
 FT DISULFID 365 374 BY SIMILARITY.
 FT DISULFID 377 395 BY SIMILARITY.
 FT DISULFID 398 411 BY SIMILARITY.
 FT DISULFID 400 426 BY SIMILARITY.
 FT DISULFID 428 437 BY SIMILARITY.
 FT DISULFID 440 455 BY SIMILARITY.
 FT DISULFID 458 472 BY SIMILARITY.
 FT DISULFID 460 479 BY SIMILARITY.
 FT DISULFID 481 490 BY SIMILARITY.
 FT DISULFID 493 507 BY SIMILARITY.
 FT DISULFID 773 785 BY SIMILARITY.
 FT DISULFID 775 792 BY SIMILARITY.
 FT DISULFID 794 803 BY SIMILARITY.
 FT DISULFID 806 818 BY SIMILARITY.
 FT DISULFID 821 833 BY SIMILARITY.
 FT DISULFID 823 840 BY SIMILARITY.
 FT DISULFID 842 851 BY SIMILARITY.
 FT DISULFID 854 864 BY SIMILARITY.
 FT DISULFID 867 876 BY SIMILARITY.

Query Match 3.2%; Score 172; DB 1; Length 1786;
Best Local Similarity 19.4%; Pred. No. 0.00017;
Matches 147; Conservative 61; Mismatches 207; Indels 344; Gaps 41;

QY 61 COE-KDYHFEYTE-----CDSSGRNRVAPNSAVDCSLPDPVRGKCTFS 106
DB 374 CSQKPFYFQHPERIDRPNLCEPCTCDPAGSE-----NGGI-CDGYTD-----FS 418

QY 107 CA--SSE---YLEMKNVCKGEGTSLGSGIKFDEWDELPAFNSIA-TFMDTVVG-- 158
DB 419 VGLIAGQCKKLUHVGRCDCVCKEGFYDLSA-----EDPYGCKSCACPLGTIPGN 470

QY 159 PSDSRP-----DGC-----NNS-----S 171
DB 471 PCSETGYCYCKRLVTGQRCDCLPQHGLSNDLDCRCDLGGALNNSCEDSGQCS 530

QY 172 WTP-----RGVIESNRDDCTVS-LIYAVHLKK--SGYVFEYQYVDNNI----- 213
DB 531 CLPHMIGROCNVESEGYFTLTDHYIYEAEEANLPGVVVVERQYIQDRIPSWTGPFGVR 590

QY 214 -----PFEFFION-DQCQEMDITTDKWKYKLTONGEWGSHSVMLKSGTNLYKRTTGILMG 267
DB 591 VPEGAYLEFFIDNIPYSMEYELIRYEPOLPDHWE----- 625

QY 268 SKAVAPVLVKNTIE---GVAYTSECFCKPG-----TFSNKPQSFNCQVCPRTYSEKA 320
DB 626 -RAV-----ITVQRPKIPASSRCGNVTNPDNDQVVSLSFGS-RYVVLPRPVCFEKG 676

QY 321 KCIRCKDQSGSGSECTE-----RPPCTTKDYFOIHPFCDDE-----GKTQI 364
DB 677 NYTVRL-ELPQYTAGSDVESPYTFIDSLVLMYPYCKSLDIFTVGGSGDGEVTSNAWETQ 735

QY 365 MYKWIE--PKICREDITDAIR-----LPPSEKEDCP-----CNGFYNNNGSS 406

Db 736 RYRCLENSRSYVVKTPMTDVCNRNIIISALIHQTLACEDPOGLSSVCDP---NGQC 792

QY 407 SCHP-----CPPGTFSGDTRECRPCPACTEPALGFYKWNVLPGNNMKTSCFNWGN 457

Db 793 QCRPNVVGRTCNRCAPGTGEGPNCCKPC-----DCHIQGS 828

QY 458 SK--CDGMNGWEVAGDHIQSGAGSDNDYLILNLHHPGFKPPTSMGTGATGSELGRITFVF 515

Db 829 ASAFCDALTG-----OCHCFQGIYARQCDRC-----LPQY----- 858

QY 516 ETLCSADCVLVFMVDINRKNVSVESMGGTKEQAYTHIIFKNATFTTTFWAFQRTNQGD 575

Db 859 -----WGFPPSCQPCQC 869

QY 576 NRRFINDMVKIYSTATNAVGVASSCRACALGSEQSSSCVCPGPHY---LEKETNQOC 632

Db 870 NHALD-----CDVTGTGECISQ--DYTTGNGERCLAGYGYDPIIGSGDHC 914

QY 633 KECPPDTYLSIHQVYVYKACIPCGPGSKNNQDHSVCYSD 671

Db 915 RQPC-----CPDGPDS--GRQFARSCYOD 936

RESULT 14

NTCL_BRARE STANDARD: PRT: 2437 AA.

ID NTCL_BRARE STANDARD: PRT: 2437 AA.

AC P46530;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Neurogenic locus notch homolog protein 1 precursor.

GN NOTCH1A OR NOTCH.

OS Brachydanio rerio (zebrafish) (Danio rerio).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;

OC Cyprinidae; Danio.

OX NCBI_TaxID=7955;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Embryo;

RX MEDLINE=94128602; PubMed=8297791;

RA Bierkamp C.; Campos-Ortega J.A.;

RT "A zebrafish homologue of the Drosophila neurogenic gene Notch and its pattern of transcription during early embryogenesis.";

RL Mech. Dev. 43:87-100(1993).

CC -!- FUNCTION: IMPLICATED IN CELL FATE SPECIFICATIONS DURING THE EMBRYO DEVELOPMENT. MAY BE INVOLVED IN THE FORMATION OF THE NEURAL PLATE, NOTOCHORD AND BRAIN VESICLES.

CC -!- SURCELLULAR LOCATION: Type I membrane protein.

CC -!- DEVELOPMENTAL STAGE: Expressed in all cells in pregastrulation stages. During gastrulation is differentially expressed, accumulating predominantly in the prechordal-mesoderm and notochord. At the end of gastrulation, expressed along the anterior-posterior axis including the developing neural plate and differentiating mesoderm. Also present in the developing brain and head regions.

CC -!- SIMILARITY: BELONGS TO THE NOTCH FAMILY.

CC -!- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.

CC -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.

CC -!- SIMILARITY: CONTAINS 6 ANK REPEATS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch)

CC EMBL; X69088; CAA48831.1;

CC HSSP; P00740; 1EDM.

DR 2FIN; ZDB-GENE-990415-173; notchla.

DR InterPro; IPR002110; ANK.

InterPro: IPR000152; Asx_hydroxyl.	FT	REPEAT	2048	2077	ANK 6.
InterPro: IPR000561; EGF-like.	FT	DOMAIN	2265	2265	POLY-GLN (OPA-REPEAT).
InterPro: IPR000742; EGF_2.	FT	SITE	1656	1657	CLEAVAGE BY FURIN-LIKE PROTEASE (BY
InterPro: IPR001881; EGF_Ca.	FT				SIMILARITY).
InterPro: IPR001438; EGF_II.	FT				BY SIMILARITY.
InterPro: IPR000800; Notch.	FT	DISULFID	25	35	BY SIMILARITY.
Pfam: PF00008; EGF_36.	FT	DISULFID	29	45	BY SIMILARITY.
Pfam: PF00023; ank; 6.	FT	DISULFID	47	56	BY SIMILARITY.
Pfam: PF00066; notch; 3.	FT	DISULFID	62	73	BY SIMILARITY.
PRINTS; PR00010; EGFBL00D.	FT	DISULFID	67	86	BY SIMILARITY.
PRINTS; PR01452; NOTCH.	FT	DISULFID	88	97	BY SIMILARITY.
SMART; SM00248; ANK; 5.	FT	DISULFID	105	116	BY SIMILARITY.
SMART; SM00179; EGF_CA; 19.	FT	DISULFID	110	126	BY SIMILARITY.
SMART; SM00001; EGF_like; 16.	FT	DISULFID	128	137	BY SIMILARITY.
SMART; SM00004; NL; 3.	FT	DISULFID	143	154	BY SIMILARITY.
PROSITE; PS50088; ANK_REPEAT; 4.	FT	DISULFID	148	163	BY SIMILARITY.
PROSITE; PS50297; ANK_REPEAT; 1.	FT	DISULFID	165	174	BY SIMILARITY.
PROSITE; PS00010; ASX_HYDROXYL; 23.	FT	DISULFID	181	194	BY SIMILARITY.
PROSITE; PS00022; EGF_1; 34.	FT	DISULFID	188	203	BY SIMILARITY.
PROSITE; PS00186; EGF_2; 28.	FT	DISULFID	205	214	BY SIMILARITY.
PROSITE; PS01187; EGF_CA; 22.	FT	DISULFID	221	232	BY SIMILARITY.
Receptor; Transcription regulation; Activator; Differentiation;	FT	DISULFID	226	242	BY SIMILARITY.
Developmental protein; Neurogenesis; Repeat; ANK repeat;	FT	DISULFID	244	253	BY SIMILARITY.
EGF-like domain; Transmembrane; Glycoprotein; Signal.	FT	DISULFID	260	271	BY SIMILARITY.
POTENTIAL.	FT	DISULFID	265	280	BY SIMILARITY.
CHAIN 1 20	FT	DISULFID	282	291	BY SIMILARITY.
DOMAIN 21 2437	FT	DISULFID	298	311	BY SIMILARITY.
DOMAIN 21 1724	FT	DISULFID	305	320	BY SIMILARITY.
TRANSMEM 1725 1747	FT	DISULFID	322	331	BY SIMILARITY.
DOMAIN 1748 2437	FT	DISULFID	338	349	BY SIMILARITY.
DOMAIN 21 57	FT	DISULFID	343	358	BY SIMILARITY.
DOMAIN 58 98	FT	DISULFID	360	369	BY SIMILARITY.
DOMAIN 101 138	FT	DISULFID	375	386	BY SIMILARITY.
DOMAIN 139 175	FT	DISULFID	380	397	BY SIMILARITY.
DOMAIN 177 215	FT	DISULFID	399	408	BY SIMILARITY.
DOMAIN 217 254	FT	DISULFID	415	428	BY SIMILARITY.
DOMAIN 256 292	FT	DISULFID	422	437	BY SIMILARITY.
DOMAIN 294 332	FT	DISULFID	439	448	BY SIMILARITY.
DOMAIN 334 370	FT	DISULFID	455	466	BY SIMILARITY.
DOMAIN 371 409	FT	DISULFID	460	475	BY SIMILARITY.
DOMAIN 411 449	FT	DISULFID	477	486	BY SIMILARITY.
DOMAIN 451 487	FT	DISULFID	493	503	BY SIMILARITY.
DOMAIN 489 524	FT	DISULFID	498	512	BY SIMILARITY.
DOMAIN 526 562	FT	DISULFID	514	523	BY SIMILARITY.
DOMAIN 564 599	FT	DISULFID	530	541	BY SIMILARITY.
DOMAIN 601 637	FT	DISULFID	535	550	BY SIMILARITY.
DOMAIN 639 674	FT	DISULFID	552	561	BY SIMILARITY.
DOMAIN 676 712	FT	DISULFID	568	578	BY SIMILARITY.
DOMAIN 714 749	FT	DISULFID	573	587	BY SIMILARITY.
DOMAIN 751 787	FT	DISULFID	589	598	BY SIMILARITY.
DOMAIN 789 825	FT	DISULFID	605	616	BY SIMILARITY.
DOMAIN 827 865	FT	DISULFID	610	625	BY SIMILARITY.
DOMAIN 867 903	FT	DISULFID	627	636	BY SIMILARITY.
DOMAIN 905 941	FT	DISULFID	643	653	BY SIMILARITY.
DOMAIN 943 979	FT	DISULFID	648	662	BY SIMILARITY.
DOMAIN 981 1017	FT	DISULFID	664	673	BY SIMILARITY.
DOMAIN 1019 1055	FT	DISULFID	680	691	BY SIMILARITY.
DOMAIN 1057 1093	FT	DISULFID	685	700	BY SIMILARITY.
DOMAIN 1095 1141	FT	DISULFID	702	711	BY SIMILARITY.
DOMAIN 1143 1179	FT	DISULFID	718	728	BY SIMILARITY.
DOMAIN 1219 1263	FT	DISULFID	723	737	BY SIMILARITY.
DOMAIN 1265 1303	FT	DISULFID	739	748	BY SIMILARITY.
DOMAIN 1305 1344	FT	DISULFID	755	766	BY SIMILARITY.
DOMAIN 1346 1382	FT	DISULFID	760	775	BY SIMILARITY.
DOMAIN 1385 1423	FT	DISULFID	777	786	BY SIMILARITY.
DOMAIN 1446 1486	FT	DISULFID	793	804	BY SIMILARITY.
REPEAT 1487 1520	FT	DISULFID	798	813	BY SIMILARITY.
REPEAT 1521 1561	FT	DISULFID	815	824	BY SIMILARITY.
REPEAT 1867 1910	FT	DISULFID	831	842	BY SIMILARITY.
REPEAT 1915 1944	FT	DISULFID	836	853	BY SIMILARITY.
REPEAT 1948 1978	FT	DISULFID	855	864	BY SIMILARITY.
REPEAT 1982 2011	FT	DISULFID	871	882	BY SIMILARITY.
REPEAT 2015 2044	FT	DISULFID	876	891	BY SIMILARITY.
REPEAT 2051 2044	FT	DISULFID	893	902	BY SIMILARITY.

FT	DISULFID	909	920	BY SIMILARITY.	
FT	DISULFID	914	929	BY SIMILARITY.	
FT	DISULFID	931	940	BY SIMILARITY.	
FT	DISULFID	947	958	BY SIMILARITY.	
FT	DISULFID	952	967	BY SIMILARITY.	
FT	DISULFID	969	978	BY SIMILARITY.	
Query Match 3.2% Score 170.5; DB 1; Length 2437;					
Best Local Similarity 19.7% Pred. No. 0.00033;					
Matches 233; Conservative 102; Mismatches 374; Indels 473; Gaps 71;					
Qy	42	QAAWAGDL-----PSSSRPLPPCQEKDYHFEYTEDSSGSRWRVAIP-----	NS 87		
Db	128	COPGWSGKTCQLADPCASN-----PCANGG-----QCSAFESHVICTCPNPFHGQTCRQD 177			
Qy	88	AVDCSGLPDPVR-GKEC-----TFSCA-----SGEYLEMKNOVC--SKCGRTYSLSGSI 134			
Db	178	VNECAVSPSPCRNGCTCINEVGLVLCRCPPEYTGPHOORLYQPCLPSPCRSG----- 229			
Qy	135	KFDEWDELPAFNSIATFMDTVVGPSPSRDPDCNNSWIP--RGNYIESNRDDCTVSL-- 190			
Db	230	-----GTCVQTSDF-----THTCSCLPGETGOTCEHNVDDCTQHACE 266			
Qy	191	-----IYAVHLKSGYVFEYQVVDNNIFFEPIQNDQCEMDT-----TTDKWVKL 237			
Db	267	NGGPCIDGINTYNCHCDK-----HWTGOYCTEDV-DECELSPNACQNGGTCHNTIGGFHCV 321			
Qy	238	TDNGEWG-----SHSVMLKSGTNIILYWR-----TTGIL----- 265			
Db	322	CVNGWTGDDCSENIDDCASACSHGATCHDRVASFCECPHGRTGLLCHLDACISNPCQ 381			
Qy	266	MGSKA-VKPVLVKNITIEGVAYT-----SECF-----PCK-PGTFNKPFGSNCQVC- 310			
Db	382	KGSNCDTPNVSKAICTCPPGYTGSACNQDIDECSLGANPCHEHGRGLNKGSGFOCKLQ 441			
Qy	311	-----PRNTYSEKAKECTR--CKDSDS-----QFSG-----SSECTERPP 343			
Db	442	GVEGPR--CEMDVNECKSNPCQNDATCLDQTGGFHCTCMPEGYEVGFQINSDDCASOPC 498			
Qy	344	CTTKDYFQI--HTPCDEBEGTQIMYKWTPEKICREDITDTRLP-----PSGEKDC 393			
Db	499	LNGKIDKINSFHCEP-----KGFSGSLCQVDVDECASTPKCKAKACTDGNPKYT 549			
Qy	394	PCNPGF-----YNGSSSCHPCPGTFSFG-----TKECRP----- 425			
Db	550	CECTPGFSIHCELDINECASS--PCHYGVCRDGVASETCDRCRGYTGRLCETNINECL 607			
Qy	426	-----CPAGTEPALGFYKWNVLPGNNKMTSC-FNVGN----- 457			
Db	608	QPCRNGTQCDRENAYICTPKGT-----TCVNCEINIDCKRKPCDY 650			
Qy	458	SKC-DCMNGWEVAGDHIQSGAGGSDN-DYLIIN-----LHIPGFKPPT 498			
Db	651	GKCIDKINGECVCEPGYSGSNINIDCALNPNCHNGTCTIDGVNSFTCLCPDGFDRAT 710			
Qy	499	SMTGAT-----GSELGRITTFVFETLCSA-----DCVLFMVDINDKSTNNVSWG 544			
Db	711	CLSQHNECSSNPCIHGSCLDQIN-SYRCVCEAGWGRNC-----DINIECLSNPCVN-GG 764			
Qy	545	THEKQAYTHIIFKNATFTTFAFORTNOGDARREFINDMVKIYSTATNAVGVASSCRA 604			
Db	765	TCKDMTSGYLCTCRAGFS-----GPNQCMNINECASNPCLNQGSCIDDVAGFKCN 814			
Qy	605	CALGSRQSGSSC-----VCPGPPGHVIEKETNOCKE-----CPPTDYLTHIQVYKGE 650			
Db	815	CML--PYTGEVCENVLAPCSPRPC--KNGVCRESEDFQSCNCPAGNGQTCCEVDINE 870			
Qy	651	ACI--PCGPGSKNQDHVSVYCDFFYHEKENOILHYDFSNSLSSVGLMNGFSFTSKGTR 708			
Db	871	-CVRNPCTNG-----GVC-----ENLRGGFQCRNPGFTG----- 899			
Qy	709	YFHFFNISLHGHEGKKMALCTNNITDFTVKEIVAGS--DDYTNLVGAFCVQSQSTIIPSEK 766			

Db	900	-----ALCENDIDCEPNPCSNNGVCQDRVN---GFVCVCL-----A 933			
Qy	767	GFRALSSQSIIADTFIGVTVETTLANIKEMDFVPTSQIPDVHFFYKSSATTTSCI 826			
Db	934	GFRGERCAEDI-----DECVSAPC-----RNGNCNCTDCV 962			
Qy	827	NGRSTAVKMRNPTKSGAGV---ISVP-----SKC-PAGTC-DGCTEYFLWESAECPLCT 877			
Db	963	NSYT-----CSPAGTSGINCEINTPDCTESSCFNGTCDVGDISSF-----SCVCLPGFT 1012			
Qy	878	-----EHDPHETE-----GACKRGFO-----ETLYVWNEPKWCIKIGSL 911			
Db	1013	GNVQHDVNECDSPRCNGSGCQDGYTYKTCYKCTPHGYTGLNCQSLVRWCDSPPCKNGSC 1072			
Qy	912	PEKKA-TCETVDFWL-----KVGAGVGAFATVLLVALTC 945			
Db	1073	WQOGASTTCQASCAGWTGIYCDVPSVSCVAAARQOGVSVAVLC 1114			
RESULT 15					
Id	LML1_CAEEL	STANDARD;	PRT;	1557	AA.
AC	Q18823;				
DT	01-NOV-1997 (Rel. 35, Created)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	01-NOV-1997 (Rel. 35, Last annotation update)				
DE	Laminin-like protein C54DL1.5 precursor.				
GN	C54DL1.5.				
OS	Caenorhabditis elegans.				
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;				
OC	Rhabditidae; Pelodierinae; Caenorhabditis.				
OX	NCBI_TaxID=6239;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=Bristol N2;				
RA	Minx P.;				
RL	Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.				
CC	-1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).				
CC	-1- SIMILARITY: CONTAINS 10.5 LAMININ EGF-LIKE DOMAINS.				
CC	-1- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/				
CC	or send an email to license@isb-sib.ch).				
CC	EMBL; U46673; AAC48152.1; .				
DR	HSP; P02468; ITLE.				
DR	WormPep; C54DL1.5; CE06981.				
DR	InterPro; IPR000561; EGF-like.				
DR	InterPro; IPR001886; LamNT.				
DR	InterPro; IPR000034; Laminin_B.				
DR	InterPro; IPR002049; Laminin_EGF.				
DR	Pfam; PF00052; laminin_B; 1.				
DR	Pfam; PF00053; laminin_EGF; 10.				
DR	Pfam; PF00055; laminin_Nterm; 1.				
DR	PRINTS; PR00011; EGF_LAMININ.				
DR	ProDom; PD002082; LamNT; 1.				
DR	ProDom; PD003031; Laminin_B; 1.				
DR	SMART; SM00180; EGF_Lam; 9.				
DR	SMART; SM00001; EGF_Like; 2.				
DR	SMART; SM00281; LamB; 1.				
DR	SMART; SM00136; LamNT; 1.				
DR	PROSITE; PS00022; EGF_1; 8.				
DR	PROSITE; PS01186; EGF_2; 1.				
DR	PROSITE; PS01248; LAMININ_TYPE_EGF; 11.				
KW	Hypothetical protein; Laminin EGF-like domain; Signal; Repeat.				
FT	SIGNAL	1	POTENTIAL.		
FT	CHAIN	?	1557	LAMININ-LIKE PROTEIN C54DL1.5.	
FT	DOMAIN	?	271	LAMININ N-TERMINAL (DOMAIN VI).	

```
FT DOMAIN 272 331 LAMININ EGF-LIKE 1.
FT DOMAIN 332 387 LAMININ EGF-LIKE 2.
FT DOMAIN 388 434 LAMININ EGF-LIKE 3.
FT DOMAIN 435 487 LAMININ EGF-LIKE 4.
FT DOMAIN 488 568 LAMININ EGF-LIKE 5 (N-TERMINAL).
FT DOMAIN 569 632 LAMININ DOMAIN IV.
FT DOMAIN 633 732 LAMININ EGF-LIKE 5 (C-TERMINAL).
FT DOMAIN 733 769 LAMININ EGF-LIKE 6 (INCOMPLETE).
FT DOMAIN 770 819 LAMININ EGF-LIKE 7.
FT DOMAIN 820 874 LAMININ EGF-LIKE 8.
FT DOMAIN 875 930 LAMININ EGF-LIKE 9.
FT DOMAIN 931 978 LAMININ EGF-LIKE 10.
FT DOMAIN 979 1025 LAMININ EGF-LIKE 11.
FT DISULFID 272 281 BY SIMILARITY.
FT DISULFID 274 295 BY SIMILARITY.
FT DISULFID 297 306 BY SIMILARITY.
FT DISULFID 309 329 BY SIMILARITY.
FT DISULFID 332 341 BY SIMILARITY.
FT DISULFID 334 357 BY SIMILARITY.
FT DISULFID 360 369 BY SIMILARITY.
FT DISULFID 372 385 BY SIMILARITY.
FT DISULFID 388 400 BY SIMILARITY.
FT DISULFID 390 406 BY SIMILARITY.
FT DISULFID 408 417 BY SIMILARITY.
FT DISULFID 420 432 BY SIMILARITY.
FT DISULFID 435 449 BY SIMILARITY.
FT DISULFID 437 456 BY SIMILARITY.
FT DISULFID 458 467 BY SIMILARITY.
FT DISULFID 470 485 BY SIMILARITY.
FT DISULFID 770 778 BY SIMILARITY.
FT DISULFID 772 788 BY SIMILARITY.
FT DISULFID 791 800 BY SIMILARITY.
FT DISULFID 803 817 BY SIMILARITY.
FT DISULFID 820 834 BY SIMILARITY.
FT DISULFID 822 841 BY SIMILARITY.
FT DISULFID 844 853 BY SIMILARITY.
FT DISULFID 856 872 BY SIMILARITY.
FT DISULFID 875 894 BY SIMILARITY.
FT DISULFID 877 901 BY SIMILARITY.
FT DISULFID 903 912 BY SIMILARITY.
FT DISULFID 915 928 BY SIMILARITY.
FT DISULFID 931 943 BY SIMILARITY.
FT DISULFID 933 950 BY SIMILARITY.
FT DISULFID 952 961 BY SIMILARITY.
FT DISULFID 964 976 BY SIMILARITY.
FT DISULFID 979 991 BY SIMILARITY.
FT DISULFID 981 998 BY SIMILARITY.
FT DISULFID 1000 1009 BY SIMILARITY.
FT DISULFID 1012 1023 BY SIMILARITY.
FT CARBOHYD 101 101 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 121 121 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 333 333 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 507 507 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 643 643 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 725 725 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 921 921 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1062 1062 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1168 1168 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1211 1211 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1244 1244 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1321 1321 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1437 1437 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1557 AA; 172723 MW; CAF0B51F8D5E8D2F CRC64;

Query Match 3.2%; Score 170; DB 1; Length 1557;
Best Local Similarity 18.1%; Pred. No. 0.0002;
Matches 173; Conservative 86; Mismatches 292; Indels 404; Gaps 49;

QY 42 COAWAGDLPSSSSRLPPCOEKDYHFEYTECDSSGSRWRVAIPN-----SAVDCSGLP 95
DB 406 COCK-----PGVTGRFCQCLDGFYDFSTNGCKNGCCGTSLLNQPRCDSSSGSCSKL 460
QY 96 DTVRGKECTFCASGEY-LEMKNQV-CSKC-----GEGTYSLGSGIKFDEWDEL 142
```

```
Db 461 N-VEGRQCD-KCKPGYFDLSTENQFCCTPCFCFHSSICNTADGYFAMNVSSVFDQDKQK 518
QY 143 PAGFSNIA-----TFMDTVVGPDSRDPGCMNSS--WIPRGNYESNRDDCTVSLIYAV 194
Db 519 WAGONRIGLQDTQWAELOKAVASDT-----DNSPYVFVAPEQFLGQDRSSYNQDLVFTL 573
QY 195 HLKSGYFEFYQYVDNNIFFEFTQNDQCEMDTTDKWKLTDNGE----- 242
Db 574 KVAK-----HVTNQDVKDIIIVGADRQELSTS-----ITAQGNFPPTTEATYRF 618
QY 243-----WGSHSVMLKSGTNILYWRIT-----GTLMSGKAVKPVLVK----- 277
Db 619 RVHADPYFGW-----YPRINELDFIGLSNITAIK---IRGTSYKDIGVL 661
QY 278---NITIEGVAVTS-----ECPKCPKPGTFSNK-----PGSFN----- 306
Db 662 SNVNLGTAGVAPSAANPKOATWIEHCLECLPGFVGOFCESSCGSFRRETRFGGPFNHCIRK 721
QY 307-----CQ-----VCPNTYSEKGAKECIRKDDSQFSGSSECTERPPCTTKDYFOI 352
Db 722 DCHNHSNCEAESGSGSCICEHNTAGDT-CERCARGYGDALQGTEDCCQKPCPDNDGPCIL 780
QY 353 HTPCDEEGKTQIMYKWIIEPKICREDLTDAILRPPSGEKKDCPPCNPFGYNNGSSSCHPCP 412
Db 781 HADGD-----VICTE-----CPNGYTGRRCDECSGDFGNP----- 811
QY 413 PGTFSDGK--BCRCPAGTEPALGFYKWMNVLPGNMKTSCFNVNGNSKCDGNGMEVAG 470
Db 812-----KDGTECEVEC-ACSGNTDPN-----SIGN--CDKITG----- 839
QY 471 DHIQSGAGSDNDYLILNLHIPGKPPPTSMTCATGSELGRITFVFETLCSADCVLVFWVD 530
Db 840-----ECKCIFNTH--GFMCENCKPGYWDAL----- 865
QY 531 INRKSTNVVSWGKTKEQAAYTHIIFKNATFTTFAFORTNQGNRRFINDMVKIYSIT 590
Db 866-----IEPKGNCQSCGF-----AAGTRRPND-----YTLL 892
QY 591 ATNAVVDG-----VASSCRACALG--SEQSGSCVPC-----PPGHYIEKETNOCKE 634
Db 893 ECNQDQGGQCDCLPNVIGIQDQCAHGFYNITSLGQCQCNCPLAGSEGTCDVNTGQC-Q 951
QY 635 CPP-----DTYLSIHQVYKCACTPCG---PGSKN-----NQDHSVYCYSDCFVHEKE 679
Db 952 CKPGVTGQRCDRCADYHFGFSANGQPCDCÉYIGSENQOCDVNSGQCIC-----RE 1002
QY 680 NQILHYDFSNLSSVGLMNGPSTSKGTYKFFHFNISLGHGEGKKMALCTNNITDFTVKE 739
Db 1003 NV-----EGRRCDQCAEN----- 1015
QY 740 IVAGSDDYTNLVGAFVCOSS--TIIPSESKGFRALSSQSIIILADTFIGVTVETTLKNNI 797
Db 1016-----RYGITOGCLPCDDCYTLQSRNVNFRKYS-----LDNTLQEIEN---- 1057
QY 798 KEDMPVPPTSQIPDVHFFYKSTATTSCINGRSTAVKMRNPTKSGAGVISVPSK 852
Db 1058-----PAPVNDTK-----FDEKVKETSRASEVWEAVKOK---TKEGGGTIKTKSK 1100
```

Search completed: May 12, 2003, 13:19:58
Job time : 37.3568 secs

GenCore version 5.1.5
 Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 12:50:23 ; Search time 41.1332 Seconds
 (without alignments)
 4823.927 Million cell updates/sec

Title: US-10-073-333A-2
 Perfect score: 5357
 Sequence: 1 MLFRANGPVRGWRGPAEA.....TCYFWKKKQKKKTLNLFN 963

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : SPTREMBL_21.*
 1: sp_archaea.*
 2: sp_bacteria.*
 3: sp_fungi.*
 4: sp_human.*
 5: sp_invertebrate.*
 6: sp_mammal.*
 7: sp_mhc.*
 8: sp_organelle.*
 9: sp_phage.*
 10: sp_plant.*
 11: sp_rodent.*
 12: sp_virus.*
 13: sp_vertebrate.*
 14: sp_unclassified.*
 15: sp_rvirus.*
 16: sp_bacteriap.*
 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2624	49.0	493	Q96DP2	Q96dp2 homo sapien
2	1557	29.1	580	Q9P2M2	Q9p2m2 homo sapien
3	1225.5	22.9	1019	Q9NA40	Q9na40 caenorhabdi
4	661.5	12.3	300	Q8R215	Q8r215 mus musculu
5	222.5	4.2	1074	Q964D1	Q964d1 entamoeba h
6	211.5	3.9	1101	Q964D2	Q964d2 entamoeba h
7	207.5	3.9	704	Q9U048	Q9u048 giardia lam
8	203	3.8	709	Q9XTJ7	Q9xtj7 giardia lam
9	201	3.8	709	Q97444	Q97444 giardia lam
10	199.5	3.7	667	Q95WU1	Q95wul giardia lam
11	197.5	3.7	3087	Q95WU1	Q95wul giardia lam
12	193.5	3.6	5374	Q99ND0	Q99nd0 mus musculu
13	192.5	3.6	719	Q9U021	Q9u021 giardia lam
14	192.5	3.6	2759	O45614	O45614 caenorhabdi
15	192.5	3.6	3102	Q9TZR4	Q9tizr4 caenorhabdi
16	191	3.6	1289	Q8SSS3	Q8sss3 dictyosteli

17	188.5	3.5	769	5	Q24971	Q24971 giardia lam
18	188.5	3.5	999	4	Q9NQ36	Q9nq36 homo sapien
19	187	3.5	1723	5	Q8WRD0	Q8wrdo plasmodium
20	186	3.5	1007	13	Q90ZK3	Q90zn3 gallus gall
21	185.5	3.5	719	5	Q9U019	Q9u019 giardia lam
22	185.5	3.5	2534	5	Q8WRD1	Q8wrdr1 plasmodium
23	184	3.4	1557	5	Q8WPK9	Q8wpk9 oikopleura
24	183	3.4	1372	5	P91526	P91526 caenorhabdi
25	182	3.4	2931	5	Q9W2C6	Q9w2c6 drosophila
26	181.5	3.4	1551	5	Q9NGV4	Q9ngv4 drosophila
27	181.5	3.4	3396	5	Q9VM55	Q9vm55 drosophila
28	181	3.4	2189	5	Q9BI05	Q9bi05 eimeria ten
29	179.5	3.4	718	5	Q9BI07	Q9bi07 entamoeba h
30	179.5	3.4	3567	11	Q9ES77	Q9es77 mus musculu
31	178.5	3.3	3704	5	P91904	P91904 caenorhabdi
32	175	3.3	2906	11	Q9WU99	Q9wuh9 rattus norv
33	174.5	3.3	548	5	Q9GQ45	Q9gg45 giardia lam
34	174.5	3.3	597	5	Q07317	Q07317 giardia int
35	173.5	3.2	468	5	Q9BMH3	Q9bmh3 ichthyophth
36	173	3.2	667	5	Q9XTK3	Q9xtk3 giardia lam
37	173	3.2	1045	5	Q8T3A6	Q8t3a6 caenorhabdi
38	173	3.2	1070	5	Q8T3A7	Q8t3a7 caenorhabdi
39	173	3.2	1111	5	Q9XWD6	Q9xwd6 caenorhabdi
40	172.5	3.2	3857	11	O88840	O88840 mus musculu
41	171.5	3.2	2809	4	Q96JP8	Q96jp8 homo sapien
42	171	3.2	1316	4	Q96JU7	Q96ju7 homo sapien
43	169.5	3.2	739	5	Q9GS24	Q9gs24 giardia lam
44	168	3.1	594	5	Q24970	Q24970 giardia lam
45	167.5	3.1	1664	5	Q9TVQ2	Q9tvq2 caenorhabdi

ALIGNMENTS

RESULT 1

Q96DP2 PRELIMINARY; PRT; 493 AA.
 AC Q96DP2;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE CDNA FLJ31340 fis, clone MESAN100035, weakly similar to major
 DE surface-labeled trophozoite antigen precursor.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_taxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
 RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
 RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
 RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
 RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
 RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,
 RA Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K.,
 RA Isogai T.,
 RT "NEDO human cDNA sequencing project."
 RL Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AK055902; BAB71041.1;
 DR InterPro; IPR001881; EGF_Ca.
 DR SMART; SM00179; EGF_CA; 1.
 SQ SEQUENCE 493 AA; 54585 MW; 0AEBB2C19709B211 CRC64;

Query Match 49.0%; Score 2624; DB 4; Length 493;
 Best Local Similarity 97.3%; Pred. No. 3.2e-215;
 Matches 470; Conservative 4; Mismatches 7; Indels 2; Gaps 1;

OY 244 GSHSVMLKSGTNILYWRRTGILMGSKAVKPVLVKNITTEGVAYTSECPCPKGTSNRPG 303
 Db 4 GNYQVMLKSGTNILYWRRTGILMGSKAVKPVLVKNITTEGVAYTSECPCPKGTSNRPG 63
 OY 304 SFNCQVCPNRTYSEKAKECIRCKDSQFS--GSSECTERPPCTTKDYFIHTPCDEEK 361

```
|||||
Db 64 SFNCQVCRNTYSEKGAKEIRCKDDSQFSEEGSSECTERPECTTKDYFQIHTPCDEBCK 123
Qy 362 TOIMYKWIPEKICREDLTDALRLPSGKDKCPCNPGFYNGSSSCHPCPGTFSOGTK 421
Db 124 TOIMYKWIPEKICREDLTDALRLPSGKDKCPCNPGFYNGSSSCHPCPGTFSOGTK 183
Qy 422 ECRPCPACTERFALGEYKWNVLPNMMKTSFNVGNSKCDGMNGWEVAGDHTQSGAGSD 481
Db 184 ECRPCPACTERFALGEYKWNVLPNMMKTSFNVGNSKCDGMNGWEVAGDHTQSGAGSD 243
Qy 482 NDYLILNLHIFGKPPPTSMGTGATGSELGRITFVFETLCSADCVLYFMVDINRKNSTNVVES 541
Db 244 NDYLILNLHIFGKPPPTSMGTGATGSELGRITFVFETLCSADCVLYFMVDINRKNSTNVVES 303
Qy 542 WGGTKEKQAYTHIIFKNATFTFWAFORTNOGODNRRTNDMWKLYSIATNAVGVASS 601
Db 304 WGGTKEKQAYTHIIFKNATFTFWAFORTNOGODNRRTNDMWKLYSIATNAVGVASS 363
Qy 602 CRACALGSEQSGSCVCPGPPGHYIEKETNOCKECPDPTLYLSHQVYKREACTPCGPGSKN 661
Db 364 CRACALGSEQSGSCVCPGPPGHYIEKETNOCKECPDPTLYLSHQVYKREACTPCGPGSKN 423
Qy 662 NODHSVCYSDCFYFHEKENQILHYDFSNSVSGSLMNGPSTSKGTGYFHFNFISLCGHE 721
Db 424 NODHSVCYSDCFYFHEKENQSLHYDFSNSVSGSLMNGPSTSKGTGYFHFNFISYVGMR 483
Qy 722 GRK 724
Db 484 GRR 486

RESULT 2
Q9P2M2
ID Q9P2M2 PRELIMINARY; PRT; 580 AA.
AC Q9P2M2;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE KIAA1324 protein (Fragment).
GN KIAA1324.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=20181126; PubMed=10718198;
RA Nagase T., Kikuno R., Ishikawa K., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.XVI.
RT The complete sequences of 150 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:65-73(2000).
DR EMBL; AB037745; BAA92562.1; -.
FT NON_TER
SQ SEQUENCE 580 AA; 63365 MW; 5E63C19265EC8E5C CRC64;

Query Match 29.1%; Score 1557; DB 4; Length 580;
Best Local Similarity 52.1%; Pred. No. 4.4e-124; Mismatches 151; Indels 8; Gaps 3;
Matches 274; Conservative 93;

Qy 440 WNNVLPNMMKTSFNVGNSKCDGMNGWEVAGDHTQSGAGSDNDYLILNLHIFGKPPPT 499
Db 1 WNNVLPNMMKTSFNVGNSKCDGMNGWEVAGDHTQSGAGSDNDYLILNLHIFGKPPPT 60
Qy 500 -MTGATGSELGRITFVFETLCSADCVLYFMVDINRKNSTNVVESWGTKEKQAYTHIIFKN 558
Db 61 VMADTENKEVARITFVFETLCSYNCELYFMVGNSRNTPTVETWKGSKGQSYTYIIEEN 120
Qy 559 ATFTFTWAFORTNOGODNRRTNDMWKLYSIATNAVGVASSCRACALGSEQSGSCVCP 618
Db 121 TTTSFTWAFORTTTFEASRKYNDVAKIYSINVTNMGVASYCRPCALEASDVGSCTS 180
```

```
Qy 619 CPPGHYIEKETNOCKECPDPTLYLSHQVYKREACTPCGPGSKNNDHSVCYSDCFYFHEK 678
Db 181 CPACVYIDRDSGTCHSCPPTILKAHOPYGVQACVPCPGTKNNKIHSCLYNDCTFSRNT 240
Qy 679 ENQILHYDFSNSVSGSLMNGPSTSKGTGYFHFNFISLCGHEGKMKALCTNNITDFTVK 738
Db 241 PTRFNFNFSAIANVTLLAGGPFSTSKGLKYFHHFTLSLGNQGRKMSVCTDNVTDLRIP 300
Qy 739 EIVAGSDDTNLVGAFCQSTIIPSEKGFRAALSSQIILADTFIGVTETTLKLNINIK 798
Db 301 E--GESGFSKSIATVCOAVLIIPVETGYKAGVSSQPVSLADRLIGVTTDTLDCITSP 357
Qy 799 EDMFPVPTSQIPDVHFFYKSSSTATTSCINGRSTAVKMRNPTKSGAGVISVPSKCPAGTC 858
Db 358 AELPHLESIGLPDVIFVYKSRNDVTQSCSSGRSTTIRVRCSPQKTPVGSLLLPCTCSDGTC 417
Qy 859 DGCIFYFLWESAECPLCTEHDHFEIEGACAKRGQETIYVWNEPKWCIGISLPEKKLAT 918
Db 418 DGCNFHFLWESAACPLCSVADYHAIIVSSCVAGTQKTTYVWREPKLCSGGISLPEQRTI 477
Qy 919 CETVDVFLKVGAGVGAFTAVLLVALTCYFWKKNNKKK---KTILN 960
Db 478 KCTIDFNLKVGISAGCTACTALLTLCYFWKKNNKLEKYKSLVNN 523

RESULT 3
Q9NA40
ID Q9NA40 PRELIMINARY; * PRT; 1019 AA.
AC Q9NA40;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE Y73F8A.5 protein.
GN Y73F8A.5
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Matthews L.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; AL132862; CAB70224.1; -.
DR InterPro; IPR002049; Laminin_EGF.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; UNKNOWN.1.
SQ SEQUENCE 1019 AA; 112378 MW; 6EB4094722B707E7 CRC64;

Query Match 22.9%; Score 1225.5; DB 5; Length 1019;
Best Local Similarity 31.5%; Pred. No. 1.9e-95;
Matches 298; Conservative 155; Mismatches 371; Indels 123; Gaps 36;

Qy 61 COEKDYHFEYTCDSGSRWRVAIPNSAVDCSLPDPVVRGKECTFSCASGEYLEMKNQVC 120
Db 25 CYDDQDYFTTYTNCDENGERWRVAVPRGGQCQCNLPTPRGLNCSCFCECHYLDLSQHC 84
Qy 121 SKCGEYVSLGSGIRFDEWDELPAFSGFNATFMDTVVGPDSRDPG-----C-N 168
Db 85 RCNCPGFFSLGGIRVEEFTVLPSEGS-----VDNWDSPDAQFSNRQSVQVECPK 135
Qy 169 NSSWIPRGWYIESNRDDCTVSLIYAVHLKSGVVFYEQYVNN--IFTEFFIONDQCE 226
Db 136 EAGWVKDGLIYIPTPCVSKLSFSANLVRPGSVFTYRMPRNRNALSQVDIRNEQCS 195
Qy 227 M-DTTTDDKVKVLT-----DNGEWGSHSVMLKSGTNNILYWRNITGILMGSKAVKPVVK 277
```

Db 196 YNDVAKSMFLKYTKKEBDEEKNQDWRKRRIELKSGANVISWIIQNMGYQASNOPIHID 255
QY 278 NTIEGVAVTSFCKPCKTFSNKGSCFNCVCPRTVYSEKGAKECIRCKDDSQSG--S 335
Db 256 RIDVLGATROCTACPPGT--SPGSAECIPCSGFSKSGGCGRC--PESQYSGFKS 313
QY 336 SECTERPPCTTKDYQIHPTPCDEEGKTQIMYKWIPEPKICREDLTDAILRPPSGEKKDCPP 395
Db 314 EXCIDRPPCRVSDYVPVREPC--TNGSSRAVYKKVLPSICRDLPSATKLPPPTPKWCTCPK 372
QY 396 CNPFGYNNSSSCHPCPPGTFSGDKTECRPCAGTEPALGFYKWNVLGPNKMTSCFNW 455
Db 373 CNPGEKNKLGCEFCFKDKDYFDG--NSCSRCPVDTPVNTYGLQYQNMVLPPLKLSTCEXI 431
QY 456 GN--SKDGMNGWEVAGDHIOGAG--GSDNDYLILNLHLPGE-----KPPSMTGA 503
Db 432 SEDVATACNIGDANIPSGDLSIASLSLEGLIAFE--DILSID--EGEWNPLAPKPSKTMK-- 487
QY 504 TGSGLGRITVFETLCSAD--CVLYFMVDIN-----RKS--TNVVESWGTKEKQAYTHII 555
Db 488 --VPVAQVTVFETSCADESCALYFIEDMSAGIKGQRESFYHFLAFAFNSGSKRVMSHTV 545
QY 556 FKNATFTFWAQT--NOGQNRREINDMVKIYSITATN-----AVDGVASSCRACALG 608
Db 546 TKNTPAFNVAFILRSGVSSGDDK--ISDETRIYAINVTNVGHRGGGQGGASQCLTCP-- 600
QY 609 SEQSGSCVPCPGHYVIEKTNCKECPDPTVLSI--HQVYGEACIPCGSGKNNODHSV 667
Db 601 HTAGGTCTPCPAGNTHVETKLCVSCPLTINVTSSRVGVKSCVPCGGLTSDNGVSC 660
QY 668 CYSDCFFYHE-----KENOILHYDES-----NLSSVGLMNGPSFTSGTKYHFHFN 715
Db 661 TAMGKIQLNOGIGGKNDSSFTYDFSPFVGRSNISGV-----RVFSREGAAYYHFFSV 714
QY 716 SLCCHEGKMAKLTNNITDFTKEIVAGSDDTNLVGAFVCSQSTIIPSESKGFRAALSSQ 775
Db 715 SLFPPNKK-----COEQDFNDFMIGIL--DODKENVEG--LACRVLTLPSSNRSKTAYVT 767
QY 776 SIILADTFIGTVETLKNINIKEDMFPVP-----TSOIPDVHFFYKS--STATTSINGRS 830
Db 768 PULLAGRLDISLNRRKHGNSLSDEVLVDSDHNTSHPLDVFVFPVSTISPACPNQ 827
QY 831 TAYKMCNPTKSGAGVISVPSKCPAGTCGCTFYFLWESAECPLCTEHDHFEIEGACKR 890
Db 828 LVVAVARCVPTKKQM--BMLPHNCPDCTGCLFVIMETAQAQPCVCSNDYETINGECEN 886
QY 891 GQETLYVWNEPKWCILKISLEPKLATCTETVDFWLKVGAGVAFTA 937
Db 887 G--KQTHSIKPKKHCVTGAASOSKEVA-----CSAFTA 918

RESULT 4

Q8R215 ID Q8R215 PRELIMINARY; PRT: 300 AA.
AC Q8R215;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 32.6 kDa protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Strausberg R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
KW EMBL; BC022655; AAH22655.1;
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 300 AA; 32642 MW; 6455109C054B6CB4 CRC64;

Query Match 12.3%; Score 661.5; DB 11; Length 300;
Best Local Similarity 49.4%; Pred. No. 4.5e-48;
Matches 117; Conservative 43; Mismatches 70; Indels 7; Gaps 2;
QY 728 CTNNITDFTVKEIVAGSDDYTNLVGAFVCSQSTIIPSESKGFRAALSSQSIILADTFIGVT 787
Db 1 CTNDVTDLRIPDGEAG--FSKSVTAYVQCVIIPSEVMGYKAGVSSQPVSLADRLVGVS 57
QY 788 VETTLKLNINIKEDMFPVPTSOIPDVHFFYKSSTATTSINGRSSTAVKMCNPTKSCAGVI 847
Db 58 TDMTLEGIVSPVLEPHETSGIPDIVFFRSNDVTFQSSSSRSTTIRLCNPMKAAPGTL 117
QY 848 SVPSKCPAGTCGCTFYFLWESAECPLCTEHDHFEIEGACKRGFQETLYVWNEPKWCTK 907
Db 118 RLPSMCSGTCGCGNFHFLWESAECPLCSASDHYTFVSSCVAGIQKTYMMREPKLCSG 177
QY 908 GTSLEPKLATCTETVDFWLKVGAGVAFTAIVLLVALTCYFWKKNQKKK----KTILN 960
Db 178 GISLPEQRVTICKTIDFWLKVGISAGCTCTAILLTCTCYFWKKNQKLEYKYSKLVN 234
RESULT 5
Q964DI ID Q964DI PRELIMINARY; PRT: 1074 AA.
AC Q964DI;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Gal/GalNAC lectin Ig12 (Fragment).
GN IG12.
OS Entamoeba histolytica.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxId=5759;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HMI:IMSS;
RX MEDLINE=21391855; PubMed=11500468;
RA Cheng X.J., Hughes M.A., Huston C.D., Loftus B., Gilchrist C.A.,
RA Lockhart L.A., Ghosh S., Miller-Sims V., Mann B.J., Petri W.A. Jr.,
RA Tachibana H.;
RT "Intermediate subunit of the Gal/GalNAC lectin of Entamoeba
histolytica is a member of a gene family containing multiple CXXC
sequence motifs.";
RL Infect. Immun. 69:5892-5898(2001).
DR EMBL: AF337951; AAK92362.1;
DR InterPro: IPR000561; EGF-like.
DR PROSITE: PS01186; EGF_2; UNKNOWN_1.
KW Lectin.
FT NON_TER 1
FT NON_TER 1074 1074
SQ SEQUENCE 1074 AA; 116782 MW; 1552E2D714EB450F CRC64;

Query Match 4.2%; Score 222.5; DB 5; Length 1074;
Best Local Similarity 19.5%; Pred. No. 7.8e-10;
Matches 194; Conservative 113; Mismatches 361; Indels 325; Gaps 52;
QY 107 CASGEVLEMKNOVCCKGEGTSLGSGIKFDEWDELPAFNSIATFMDTVWGSDSRPDG 166
Db 13 CAS-----VSGACASCDEG-YELKT-----ESGSGSTOKCTLUKETCSAFSYDG 58
QY 167 CANSWIPRGNTIESNRDDCTVSLIYAVHLKKSGYVFEYQVVDNNIFFEFTI--QNDQC 224
Db 59 SDSNS--PKCVYCENGKESDTSNNNEKCKCKNG-----VDT---CESCLSKNDKC 105
QY 225 QE-----MDTTTKVVKLTNGEWGSHS-----VMLKSGTNILYWRRTGI--LMGSKAVKP 273
Db 106 GECVIGMSTTTNGGQKLCDTVTTDEHAENCVLGTAKDSSSKQCKDKCFMYSLQSGQCTK- 164
QY 274 VLVKNITIEG--VAYTSECFPCPKPGTFSNKPQSFNCOVPRNTYSEKGAKECIRCKDDSQ 331
Db 165 ---KNEIKIEKILQVLESSCNQCADGYIINT--EKKCTKYPDHC--SKMNSDKNGCMGEYY 218
QY 332 FSGSSECTERPPCT---TKDYFOIHTPCDEEGKTQIMYKWIPEPKICREDT--DAI---- 382

```
Db 219 LMG-----TECKVCTTDSNKL-----SEGNQSIYNAEHCECNKRCRTVSDGVCVKN 266
Qy 383 --RLPPSGEKDPCPCPNPGFYNGSSSCHP-----CPPGTF--SDGTKE--CR 424
Db 267 HCLRLFSPTENKCTKCDGQYFLTGAGKCSPLNDGFKTSKTECKQGYLYEKDGDKRCS 326
Qy 425 PCPAGTEPALGFYKWNVLPGNMK-----TSCFNWGNK 459
Db 327 LCPDPFTECLTSQ-----TPVPGKLNRLSAHLNSTDGPCKPLGCLLCSDDDTICY-----K 377
Qy 460 CDGMNGWEVAGDH-----IQSGAGSDNDYLILNL-----HIPGRKPPPTSMGTGTS 506
Db 378 CE--NGLTLNGTHCHYCNFDVKVKGVLGTSGNHQVCKMRGVDQYQYLNAFKASDN-----428
Qy 507 ELGRITFVETLCS--ADCVLYFMVINDKSTNVVSGWGTEKQKQYTHIIFKNATFTFTW 565
Db 429 ----TY-----YCLPLKDLPLPYFYSVTGSDNKKITIGCVGKDRDV-----464
Qy 566 AFQRTNQGQDNRRFI-----NDMKYIYSITATNAVGVASS--CRACALGSESGSGSC 616
Db 465 ----KNDCECDKYPKSVKASDCVSIK--TKLPSCEAANENICTQCPVSGSHVDSNGK 518
Qy 617 VCPPPGHYTEKETNOCKECPD--TYLSIHQVYKCAIPC-----GPGSKNQDHSVCYS 670
Db 519 CCGDAHYFD--QNNKQCECPASCSCSYDSSKSVVCSYECYENIOGVSTROKDNCEACK 577
Qy 671 DCGFFYHEKENQILHYDFSLSVGLMNGPSFTSKGTIFYHFFNISLSCGHEG-----722
Db 578 DTPYKEGLNA-----EDKKKSCAQLNNN-----CKEEGHYKISDGF 614
Qy 723 -----KMAKLTNNITDFTVKEIVAGSDDYTNLVGAPVCOSTIIPSES 765
Db 615 ITCLEDDSAIYDSQTECAQCASN-----AFKDNNK-----CQLCSTKKDK 658
Qy 766 KGFRAALSOS--IILADTFIGVTVTETLKNINIKEDMPVPVTSQIPDVHFFYKSTATTS 824
Db 659 YGHCASACATACIEDTNLVLAAAGSNQAQCTVCKDGFQIESPTDGY-----CSPCPAK 714
Qy 825 CINGRSTAVK--MRCNPTKSGAGVISVPSKCPA-----GTCDGCT-----862
Db 715 CKTKYSADKKEIECTCTDQSSVDIKPTCACTGTVQLENGTSCSDLSKYPGCKTT 774
Qy 863 -----FVLESFA-----CPLCTEHDFEIEGACKRGFQETLYVWNEPK 903
Db 775 DTCNVDSRTGIYATCECDGFGSRPSYNTCTCIESNYYPKEG-----EKN 820
Qy 904 WCIKGISLPEKKLAYCETDWF-----LKVGA 930
Db 821 GCAK-----CDDKCATCSDKDCTCLTCTDPLKGS 849

RESULT 6
Q964D2
ID Q964D2 PRELIMINARY; PRT: 1101 AA.
AC Q964D2:
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Gal/GalNac lectin Igl1.
GN IGL1.
OS Entamoeba histolytica.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=5759;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HMI:IMSS;
RX MEDLINE=21391855; PubMed=11500468;
RA Cheng X.J., Hughes M.A., Huston C.D., Loftus B., Gilchrist C.A.,
RA Lockhart L.A., Ghosh S., Miller-Sims V., Mann B.J., Petri W.A. Jr.,
RA Tachibana H.;
RT "Intermediate Subunit of the Gal/GalNac Lectin of Entamoeba
histolytica Is a Member of a Gene Family Containing Multiple CXXC
```

```
RT Sequence Motifs ";
RL Infect. Immun. 69:5892-5898 (2001).
DR EMBL; AF37950; AAK92361.1; -
DR InterPro; IPR000561; EGF-like.
DR PROSITE; PS01186; EGF_2; UNKNOWN_1.
KW Lectin.
SQ SEQUENCE 1101 AA; 119512 MW; C8B6F5CBD656AEC CRC64;

Query Match 3.9%; Score 211.5; DB 5; Length 1101;
Best Local Similarity 20.0%; Pred. NO. 7e-09;
Matches 202; Conservative 104; Mismatches 351; Indels 355; Gaps 59;

Qy 101 KECTSCASGEYLEMKNQVCSKCGEGTYSLGSGIKFDEWDELPAQFSNIATFMDTVVGPS 160
Db 27 REAVPHCAS-----VSGACTSCDTG-YELTT-----TCNNKTCTLLKEDMCKTA 69
Qy 161 DSRPDGCHNSSWIPRGNYI-----ESNRDCTVSLIYAVHLKKSQYVFFEQVVD 210
Db 70 FSYDYKTNSTN--PKCTYCVNGKEVNTSSHSNDKCVCK-----106
Qy 211 NNIF--EFTQNDQOE---MDTTDKWVKLPDNGEMGSHV---MLKSGTNI---256
Db 107 NNVNICESCLLMKDSKCECIIGMSTTVTD-GSKLCDNATTEDHAENVCVGLLASSTSSKTC 165
Qy 257 ----LYWRTTICILMGSAKVPVLVKNITIEG--VAYTSECPPCKPCTFSNKPGEFNCQV 309
Db 166 DKCFGM-----SLOGGKCTQ-----KNDKINKCILQVENSQCQADG-YSLSTDKKSKNK 215
Qy 310 CPRNTYSEKAGEKTRCKDDSOFGS-SECTERPCTTKDYFIHTPCD-EEGKTQIMYK 367
Db 216 FFEHC-SKINGNOCLTCMEGYLSKTDKCT---ICT-----VDNPNLSENGECSIYN 265
Qy 368 WIEPKICREDLT--DAI-----RLPPSGEKKDCPPCNPFGYNNSSSCHP-----410
Db 266 ABHCTSCNKRCTVSGVCVKHCRLFSPTEENKCTKCDGVELTTSCTCSPLNYDGFKTA 325
Qy 411 ----CPPGTF--SDGTKE--CRPCPAGTEPALGFYKWNVLPGNMK-----449
Db 326 NRTECENGYYLEKDKKRCLCPDPFTTECLTSK---TPVPGKLNLRSSHLTSTDGPK 381
Qy 450 -----TSCFNVGNSKCDGMNGWEVAGDH-----IQSGAGSDNDYLILNL-----489
Db 382 LPGLCLLCSDDDTICY-----KCE--NGLTLNGTHCHYCNFDTSKVLCTSGNNHQVCKMRGVD 434
Qy 490 ----HIPGFKPPTSMGTGATSELGRITFVFETPLCSADCVL-----YFMVDINKRSTNV 538
Db 435 QYEQLNAFAKASDN-----TWY-----CPLKDLPLPYFYSVT-----466
Qy 539 VESWGTKEKQAYTHIIFKNATFTFTWAFORTNOGOD---NRREI-----NDMKIYS 588
Db 467 ----KGTSN-----TITIGCVGQLHNVSNDCECDKHIPTSIDKASDCVSTTT 511
Qy 589 ITATNAVDCVASSCRACALGSGSSGSCVPCPPGHVIEKETNOCKECPD--TYLSIHQVY 647
Db 512 KLPSCERTANGNICTQCPVGSIVHGDKGKSCGDAHYFDKD-NVCKKCPASCSSCYDSSK 570
Qy 648 GKEACIPC-----GPGSKNNQDHSVCYSDCFYFHEKKNQILHYDFSNSLSSVGLMNGPSF 702
Db 571 SKVVGSECVENIQGVTRTKNECACINDGY---KE-----GPN 607
Qy 703 TSKGTKYFHFHFNISLCGHEGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAPV-----CQ 757
Db 608 EDKKKSCAQLNN--NCKKEGK-----YEISDGFV--TCLDCCDSAYIVGSOVGAQTQS 657
Qy 758 STIIPSESKGFRALSSQSIIILADTFIGVTETTLKNINIKEDMPVPVTSQIPDV---HF 814
Db 658 PNAFKDENNKCOLCSTKQSOYGHCAACASATACITCEDINL-----ILTGEKPCVCKDG 711
Qy 815 FYKSTAT-----TSC--INGRSTAVKMRK--NPTKSGAGVISVPS-KCPA-----855
Db 712 FYQIENATDGVYCSPCPAKCKTKYNTTSKKVECVTCTEQRLKDIKAPECACPTGTVQLE 771
Qy 856 -GTCDGCT-----FYFLWESAEA-----CPLCTEHDFHEI 884
```

```
Db 772 NGTCQSCDLSKYPGCKTKDSCNVDSTRTGFIATGECSDGFSGRSPYSNCTTCTKSNYPK 831
QY 885 EGACKRGFOETLYVWNEPKWIKGISLPEKKLATCETVDFW-----LKGVA 930
Db 832 EG-----EKNGCAK-----CDKCATCSDKDKTCLTCAADPLKVG 865

RESULT 7
Q9U048
ID Q9U048 PRELIMINARY; PRT; 704 AA.
AC Q9U048;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DE Variant-specific surface protein.
GN VSP417-6.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardinales; Giardia.
OX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AD-1;
RA Ey P.L., Darby J.M.;
RT "A new member of the vsp417 subfamily of variant-specific surface
protein (VSP) genes in Giardia intestinalis."
RL Submitted (May-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF065606; AAF02907.1;
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF03302; VSP; 2.
DR SMART; SM00261; FU; 4.
SQ SEQUENCE 704 AA; 71677 MW; 7E5AE1245AD4FD45 CRC64;

Query Match 3.9%; Score 207.5; DB 5; Length 704;
Best Local Similarity 20.8%; Pred. No. 8.2e-09;
Matches 199; Conservative 88; Mismatches 353; Indels 315; Gaps 54;

QY 38 ALAQCOAWAGDLPSSSRPLPPCQEKDYHFEYTECDSSGRWRVAIPNSAVDCSGLPD 96
Db 15 ARAAQADGSGDCKTCGVITGE-----QYCSECNGA-----NYA----- 52
QY 97 PVRGKECTFSCASGEYLEW---KNQ-VCSKCG-----EGTSLGSGIKFDEWDELPA 145
Db 53 PVNG-QCADVNAEGOSKTLCPAKGCKTQCAGSFMFKDCYS----- 95
QY 146 FSNIAFTMDTVVGPS---DSRPDGCNNSWIPRGNYIESNRDDCTVSLIYAVHLKSGV 202
Db 96 -----KDTAPGQSMCTQASEGKCTEAAPGYFLNPLRANTKDSVSCSDTTGFTDSGKT 148
QY 203 FFEYOYVNNIFFEFFIIONDOCEQMDTTTDKWKLTDDNGEMGSHVMLKSGTNIILYR 262
Db 149 YRGVQH-----CERCDGAA-----LTDA--GGDAKCTRCGD-KYLATT 185
QY 263 GILMGSKAVKPLVKNITIEGVAITSECFPCPKPGTFSKNPGSFN---COVCPRNYSSE 319
Db 186 GTCG-----EG-----CTPDTFESKEDSDNGKRCFACGDVT---TG 218
QY 320 AKECIRCKDDSQFSGSSSECTERPPCTTKDYIHTPCDEEGKTOIMYKIEPKICREDLT 379
Db 219 VASCEKCTPPDQAKPACTK---CGGNYY--LTKAAD--GTTTC-----VQSACSPD-S 266
QY 380 DAIRLPPSGEKKDPPCNPNGYNNSSSCHPPPTGSDGTEKRCPCPAGTEPALGFEYK 439
Db 267 FVENSSQSGNR--CVLCGDA-ANGVDKCAACTP-----TDQRIAPTITCIACNNGY 317
QY 440 WNNVLPGNWKTCFNVGNSKCDGNGWEVAGDHTQSGAGGSDNDYLIILNHLIPGFKPPTS 499
Db 318 -----PSADKTTCEAVSNCKTPGCKACNEGKENEVCVTCDCDSSTYL-----TPTS 362
QY 500 MTGATGSELGRITVFETLCSADCVLVPMVDINKRSTNVVSWGTEKEQAYTHLPKNA 559
Db 363 Q-----CIDSCA-----KIGNYGATEGAKK-----ICKEC 388
```

```
QY 560 TPTFTWAFORTNOGQDNRFRINDMVKIYSITATNAVGVASSCRACALGSESGSCVPC 619
Db 389 TA-----ANCKTCDGQ-----GQCACSDGFKYKNGDACSPC 419
QY 620 PRGHYTEK-----ETNOCKECPDPTVLSIHQVYKREACI-PCGPGSKNNOHDSVCYSDCF 673
Db 420 ---HESKCTCSAGTASDTECTCTGKALR-----YGDGTGKTGCEGCTTGQSGSAC----- 467
QY 674 FYHEKENQILHYDFSNLSSVGLMNGSPFTSKGTRYFHFFNISLQCGHEGKRMALCTNNIT 733
Db 468 -----KTCGLTIDGASYCSECAITTEYPQNGVC---APKASRATPTCN 507
QY 734 DFTVKEIVAG---SDDYTNLVGAFVQCSTIIPSESGFRAALSSQSILILADTIGVTVEPT 791
Db 508 DSPIONGCGTCADNYFKMNGG--CYETV---KYPGKTVTCISAPN-----GGTCQKA 554
QY 792 LKNINIKEDMPVPPTSQIPDVHFFYKSSSTATTSCTNG--RSTAVKMRCN---PTKSGAGV 846
Db 555 ADGYKLDSTLTVCSEGCKEC-----TSSTDCTTCLDGYVKSASACTKCSSCETCNGA-- 608
QY 847 ISVPSKCPAGTCDGC-TFYFLWESAE-ACPLCTEHDFHIEGACKRGFOETLYVWNEPKW 904
Db 609 -----ATTCKACATGYKTAGGEGACTSC-ESDSNGVTGI-----KG 644
QY 905 CIKGISLPEKKLAT--CETVDFWLKVGAGVGAFTAVL-----LVALTCYFW 948
Db 645 CL-NCAPSSSTGSLVCLYLQNTNKSGLSAGAIAGISAVIVVAVAGLGLCWFF 698

RESULT 8
Q9XTJ7
ID Q9XTJ7 PRELIMINARY; PRT; 709 AA.
AC Q9XTJ7;
DT 01-NOV-1999 (TremBLrel. 12, Created)
DT 01-NOV-1999 (TremBLrel. 12, Last sequence update)
DE Variant-specific surface protein (Fragment).
GN VSP417-6.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardinales; Giardia.
OX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRIS-136;
RX MEDLINE=95026095; PubMed=9806870;
RA Ey P.L., Darby J.M.;
RT "Giardia intestinalis: conservation of the variant-specific surface
protein VSP417-1 (TSA417) and identification of a divergent homologue
encoded at a duplicated locus in genetic group II isolates."
RL Exp. Parasitol. 90:250-261(1998).
DR EMBL; U89266; AAD03483.1;
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR005127; Giardia_VSP.
DR Pfam; PF03302; VSP; 2.
DR PRINTS; PR00290; KAZALINHBTR.
DR SMART; SM00261; FU; 3.
FT NON_TER 709
SQ SEQUENCE 709 AA; 71516 MW; 3512BB844B38D134 CRC64;

Query Match 3.8%; Score 203; DB 5; Length 709;
Best Local Similarity 20.8%; Pred. No. 2e-08;
Matches 162; Conservative 70; Mismatches 272; Indels 273; Gaps 40;

QY 38 ALAQCOAWAGDLPSSSRPLPPCQEKDYHFEYTECDSSGRWRVAIPNS--AVDCSGLP 95
Db 15 ARAAQADGSGSAGS-----CKTCG-----VAIQEQYCSCECNGAN 50
QY 96 -DPVRGKECTFSCASGEYLEM---KNQ-VCSKCGEYVSLGSGIKFDEWDELPAFSGFNA 150
Db 51 YAPVNG-QCADVNAEGPSKTLCPAKDQKCTQCAGNSFMLAGGC-YSSGEGLPGSHSLCS 108
```

```
QY 151 TFMVTVPVGPDSRDPDCCNNSSWIPRGNYIESNRDDCTVSLIYAVHLKKSGVVFYEQY-- 208
Db 109 SGGDGV-----CTEAAPGFLNPLRANTDSVVSCSDTAGFTDGTGTYRGVQVCE 158
QY 209 -VDNNIFE-----FFIONDOCOEMDTTIDKWKL-TDNGE--WGSHSVY 249
Db 159 RCDGAVLDAAGDAKTRCGENKYLATGTGCGEGCTPDTEFSKEDSDNGKRCFACGDV- 217
QY 250 LKSGTNILYWRITGTLMSKA-----VKPLVK-----NITIEGVAVTSECFPCPK 295
Db 218 -----TTGVASCEKCTPPSPDOAKPACTKCGNNYLNKTAADTTTCAEOSACSP 266
QY 296 CTF--SNKPGSFNCQV-----CPRNTYSEK-----AKECIRCKDDSQFSG----- 334
Db 267 DSFPVENSQGNRCVLCGDAANGVDKCAACTPADGRAAPAVTCTACTDGYKPSADTKT 326
QY 335 -----SSECTERPPCTTKDYQIHTP-----CDEEKTQIMYKWB--PKI 373
Db 327 CBVSSCKTPGCKACSGNEKENEVCTDCDGYLTPTSQIDSCAKVGNYYGAIEGAKKL 386
QY 374 CREDITLDAIRLPPSGEKKDPCPNFNGSS-----SCHPCPPGTFSOGTKECRCP 427
Db 387 CKE-CTAANKCTCDQGR-CQTCSDGFGKNGDACSPCHESCKTCSAGTASD-----CTECP 440
QY 428 ACTEPALGFYKWNVLPENMKTSFNVGNSKDCGMNGWEVAG----- 470
Db 441 TKGALRYGDD-----GTRGTC-GAGCATGCGSGACKTCGLTIDGASYGSECATATE 490
QY 471 -----DHIQSGAGGS-DNDYLIIN-----LHIPG-----FKPP-- 497
Db 491 YPONGVCAPKASRATPTCNDSPIONGVCGTCANSYFKMNGGCVETVKYPGKTVCTISAPNG 550
QY 498 -TSMTGATSELGRITFVPELTCLSDCVLYEMVDINRKSTNVVSW-----GOTKEKQAVT 552
Db 551 GTCQKAAQYKLDSTL--TVSEGC-----KECTSTDTCTCLDGVVKSASACT 598
QY 553 HIFKNATFTTFAFORTNOGQDNRFRINDMYKIYSITATNAVGVASSCRACALGSEOS 612
Db 599 KCDFSCT-----CNGAATTCACATGYKTT 624
QY 613 GSCVPCPGHYIEKETNOCKECPDPTLYSIHOVYGKEACIPCGSGKNNQDHSVCY 669
Db 625 ASG-----EGACTSCESD-----NGVTGKGLNCAPPSS-SNTGSVLCY 663

RESULT 9
O97444
ID O97444 PRELIMINARY; PRT: 709 AA.
AC O97444;
DT 01-MAY-1999 (TremBLrel. 10, Created)
DT 01-MAY-1999 (TremBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Variant-specific surface protein 417-4 (Variant-specific surface
DE protein type 4 TSP11/TS417-like).
GN VSP417-4.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
OX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AD-1;
RA Ey P.L., Darby J.M., Mayrhofer G.;
RT "A new locus (vsp417-4/A-I) belonging to a subfamily of ts417-like
RT variant-specific surface protein (VSP) genes in Giardia
RT intestinalis."
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 476-705 FROM N.A.
RC STRAIN=VARIOUS STRAINS;
RA Ey P.L., Darby J.M.;
RT "Identification of a subset of ts417-like genes within the variant-
RT specific surface protein (VSP) gene family of Giardia intestinalis."
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
```

```
DR EMBL: AF065168; AAD04339.1; -.
DR EMBL: U89153; AAD05041.1; -.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR005127; Giardia_vsp.
DR Pfam: PF03302; VSP; 2.
DR SMART: SM00181; EGF; 1.
DR SMART: SM00001; EGF-like; 3.
DR SMART: SM00261; FU; 5.
SQ SEQUENCE 709 AA; 72664 MW; 6E71F27D2F367F52 CRC64;

Query Match 3.8%; Score 201; DB 5; Length 709;
Best Local Similarity 19.6%; Pred. No. 3e-08;
Matches 172; Conservative 79; Mismatches 297; Indels 328; Gaps 45:

QY 91 CSGLPDPVRGKECTFSCASGEYLEMKNOVSKCGEGTYSLGSGIKFD---EWEDELPAQFS 147
Db 19 CTQEAADCKCKTCGVTIGOEQY-----CSEGCANYPVNGACEDVETQADKALCKA 71
QY 148 NIATFMDTVVGPDSRDPDCCNNSSWIPRG-----NYTESNRDD 185
Db 72 HASGACTTCGGNSFMFKDCYSSGEGLPGLSLSSDGDGVCYTEAAPGYFLNPLRANTKD 131
QY 186 CTVSLIYAVHLKKSGVVFYEQYVVDNNIFFEYFQNDQCOEMDTTT----- 231
Db 132 SVVSCSDTPTGFTDSGKTYRGVOY-----CERCDGAALTDAAAGGDAKCTRC 176
QY 232 --DKWVK---LTDNGEWGSHSVMLKSGTNILYWRITGILMSKAVKPVLVKNITIEGVAY 286
Db 177 GQDKYLKDKNTCVKQAQDSGS-----TN-----KFAVAVDSEN 209
QY 287 TSECPCKPGTFSNKPFGSFNCVCPRNHYSEKAK-ECIRCKDDSQFSGSSE---CTERP 342
Db 210 GNKVCVSCDNL---NGGVANDTC---SYDEQSKKIKCTKCTDNNLYLTTSEGTSCVOKD 263
QY 343 PCTTKDYQIHTPCDEEGKTOIMYKWBPKICREDLTDAILRPPSGEKKDCCPCNPGFVN 402
Db 264 QC--KDG----FPKDDSS-----AGNK--CLPCNDS--T 288
QY 403 NGSSSCHPCP--PGTFSOGTKECRCPAGTEPALGFYKWNVLPENMKTSFNVGNSKC 460
Db 289 DGIANCATCALVSGRGAALVTCASCTDGYKPSAD-----KTTCEAVSNCKT 335
QY 461 DGMNWEVAGDHIQSGAGSDNDYLIINLHIPGFKPPTSMTCATGSELGRITFVETLCS 520
Db 336 PGCKACSGNEKENEVCTDCDSSTYL-----TPTSQ-----CI 367
QY 521 ADCVLVFMVDINRKSTNVVSWGSGTKKQAYTHIIFKNATFTTFAFORTNOGQDNRFP 580
Db 368 DSCA-----KIGNYVGATEGAKK-----ICKECTA---ANCKTCDGQ----- 401
QY 581 NDWKIYSITATNAVGVASSCRACALGSEQSGSCVCPGPHYIEK-----ETNOCKEC 635
Db 402 -----GQCQACSDGFYKNGDACSPC---HESCKTCSAGTASDCTEC 439
QY 636 PPDYLSIHQVYGKEACI-PCPGSKNNQDHSVCYDCFFYHEKENQILHLYDFSNLSSVG 694
Db 440 PTGKALR-----YGDGDTGKTGCGAGCATGAGAC-----KTCG 473
QY 695 SLMNGPSFTSK---GPKYHFFFNISLCGHEGKKMALCTNNITDFTVYKEIV---AGSDDYT 748
Db 474 LTIIDGASYCSECATATEY-----PQNGVCSST---TVRAAATCKAGS----- 512
QY 749 NLVGAFVCGOSTIIPSEKGFRAALSSQSIILADTFIGTVETT-LKNINIKEDMPFV-PT 806
Db 513 --VAKGMCMSC-----TNGFL-----RMNGGCVETTKFPCKNVCEAAPAGDT 553
QY 807 SQIPDVHFFYKSSATTTCINGRSTAVKM-RCNPTKSGAGVIS-----VP-----SKPC 854
Db 554 CQTP-ADGKXNLNNGALITCSAGCKTCTSDQCDCTCAGTAKTGCNTKCVPCATGCGSEC 612
QY 855 AGTCDGCTF-----YFLWESAECPLCTEHDFHEIEG 886
```

Db 613 ADDATKTVCAAGYLL--SKEKCIACDKSDGGSITG 646

RESULT 10
Q95WU1

ID Q95WU1 PRELIMINARY; PRT; 667 AA.

AC Q95WU1; (Created)

DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)

DE Variant-specific surface protein WP/9510-B.

OS Giardia lamblia (Giardia intestinalis).

OC Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.

OX NCBI_TaxID=5741;

RN [1]

RP SEQUENCE FROM N.A.

RA Lujan H.D.;

RT Identification of a novel variant-specific surface protein of Giardia intestinalis.

RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.

DR EMBL; AF293416; AAK97086.1; .

DR InterPro; IPR000561; EGF-like.

DR InterPro; IPR005127; Giardia_vsp.

DR Pfam; PF03302; VSP; 2.

DR PROSITE; PS01186; EGF_2; UNKNOWN_1.

DR PROSITE; PS01186; EGF_2; UNKNOWN_1.

SQ SEQUENCE 667 AA; 68451 MW; 7693720320689546 CRC64;

Query Match 3.7%; Score 199.5; DB 5; Length 667;
Best Local Similarity 19.6%; Pred. No. 3.7e-08;
Matches 149; Conservative 70; Mismatches 275; Indels 265; Gaps 39;

QY 37 WALAG--QQAAGADLPSSSRPLPPCQKDYHFEYTCDDSSGRWRVAIPNSAVDCSGL 94

Db 5 FVLAVLVQIAWAGKATERAAQ---CADN-----TNCAEEACNLVIGGLKLYCSRNTG 54

QY 95 PDVVRG---KE-CTSCASGEYLEMKNQVCSKCGEGTSLSGIKFDEWDELPAFMSI 149

Db 55 FVPINGQCADKEATDQCKDGGSDTADQTCQCAEQTEMVYKGGCY--BAAQPG----- 107

QY 150 ATFMDDTVVGPSPRDPCCNNSWIPRGNIYESNRDDCTVSLIYAVHLKSGYVFEYQV 209

Db 108 ----QTMCAADA--GVCTQAA--QGIVFPACADASHQSVIPCG--DEEGIT-----V 150

QY 210 DNNIFEFFIQQDOCEMDTTDKWVKLTDNGENGSHVLMKSGTNILYWRRTTGILMSGK 269

Db 151 KNDKKYKGLHCTRCYAPTEAADAANAATACTACGDSKIV-----K 191

QY 270 AVKPVLVKNITIEGVAYTSECFPCPGTF-----SNKPGSFNCVCPRTYSEKGAKE 322

Db 192 TAKDSATSCVTEECTGTCTCAEGTSDGCATCEKGADGAVACKTCSNKKVQPNKKG 251

QY 323 CI-RCKD--DSQFSGSSECTE-----RPPCTTKDYFOIHTP-----CDECKTQIMYKW 368

Db 252 CIAKCPETVSAEKDGVCEGVVDPNAGTGCTKKPDPCNTPGCKTCS-----300

QY 369 IEPKICREDLT--DAIRLPPSGE-----KK-----391

Db 301 -EPKTSKEVCTCEDPQKALPTGQCIYGCHELEGYEGTSEGGKACKKEVENCLLCNG 359

QY 392 --DCPPCNPFYNNG-----SSSCHPCPPGTFSDGTRECRP-----CPAGT 430

Db 360 QQCETCKDGYKSGAACKNTSCKTCAANGN-SNGCTSCPEKQVLSVEGEGTGCKPKC 418

QY 431 EPALGFYKWNVLPGNMKTSCFNV-GNSKCDGMNGWEVAGDHIQSG-----476

Db 419 KPVSGK-----DGTCKSCDLNIDGTSYCSACN--VGTEYPENGVCVKKSARTASCO 468

QY 477 -----AGGSNDYLIIN-----LHPTG-----FKPPTSMGTATGSELGRITFV 515

Db 469 AEPNSGVCGTCARGFRMNGGCVETTKLPFGKSVCEVASAGDTCTPADGYKLNGALI- 527

QY 516 ETLCSDACVLYPMVDINRKSTNVVESWGTREKQAYTHLIFKNATFTTWFAPORTNOQOD 575

Db 528 --TCSAGC-----KTCTSOD 540

QY 576 NRRFINDMWKI-YSTITATNAVGVASSCRACALGSEQ-----SGSSCVPCPPGHYIEKETN 630

Db 541 Q-----CDTKAGYAKTGGN-----TKKCVPCATGCGSECNADDAKCTVCAAGYLSKE-- 589

QY 631 QCKECPDPTVLSIHQVYKREACIPCGSGKKNODHSVCY 669

Db 590 KCIACDKSGGSI---TGVANCAACAPPT--NNKGPVLCY 624

RESULT 11
Q8WRD2

ID Q8WRD2 PRELIMINARY; PRT; 3087 AA.

AC Q8WRD2; (Created)

DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)

DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)

DE Cysteine repeat modular protein 1 PBCRM1.

OS Plasmodium berghei (strain Anka).

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI_TaxID=5823;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-ANKA;

RA Thompson J.;

RT "Plasmodium berghei Cysteine Repeat Modular Protein 1: PBCRM1.";

RL Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.

DR EMBL; AF449196; AAL47156.1; .

DR InterPro; IPR000345; CytC_heme_bind.

DR InterPro; IPR000561; EGF-like.

DR InterPro; IPR001664; IF.

DR InterPro; IPR000001; Kringle.

DR InterPro; IPR000589; Ribosomal_S15.

DR Pfam; PF00051; kringle; 1.

DR SMART; SM00130; KR; 1.

DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.

DR PROSITE; PS00186; EGF_2; UNKNOWN_3.

DR PROSITE; PS00226; IF; UNKNOWN_1.

DR PROSITE; PS00070; KRINGLE_2; 1.

DR PROSITE; PS00362; RIBOSOMAL_S15; UNKNOWN_1.

DR PROSITE; PS00362; RIBOSOMAL_S15; UNKNOWN_1.

SQ SEQUENCE 3087 AA; 362575 MW; 38BDD25AB65FD43D CRC64;

Query Match 3.7%; Score 197.5; DB 5; Length 3087;
Best Local Similarity 20.0%; Pred. No. 4.7e-07;
Matches 197; Conservative 127; Mismatches 322; Indels 337; Gaps 55;

QY 102 ECTFSCAGEYLEMKNQVCSKCGEGTSLSGIKFDEWDELPAFMSIATFMDTVVGPSPD 161

Db 1317 EFSIACEDGHY--HKHKKCHPCCKGYNNLNK-----QNN 1351

QY 162 SRPDGC---NNSSWIPRGNIYESNRDDCTVSLIYAVHLK-----SGYVFEYQY 208

Db 1352 ERYKKTACGQNRRTLDKKEFEKN---CLCDLGYE-YIKNPNPNRNFCTSCPCVGEYKD 1407

QY 209 VDNNIFEFFIQQDOCEMDTTDKWVKLTDNGENGSHVLMKSGTNILYWRRTTGILMSG 268

Db 1408 V-----ISNELCK-----GIVCKKNASYL-----ILDK 1430

QY 269 KAVKPVLVKNITIEGVAYTSECFPCPGTF--SNKPGSFNCVCPRTYSEKGAKECIRC 326

Db 1431 KSVDP-----SQCL-CNGGYLYRYDKNNNEICTKINNHY-----C 1465

QY 327 KDSQFSGSSECTE-----TERPPTTKDYFOIHTPDEEGKTOIMYKWPCKICRE----- 376

Db 1466 PNDNY--YKCKPIHKTEIEIKRT--DFETINSCLECEGYEPINNMKIKDTSRDYHYNN 1521

QY 377 -----DLTADIRLPPSGEKKDCPPCNPFYNNGSSS--CHPCPPG--TFSDGTK---E 422

Db 1522 IFVTKYKNLSYIV-----NDNNICHECNLGFYKNNISSEKCIKCPRSSSTTTTYSKSTON 1576

QY 423 CRPCPAGTEPALGFYKWNVLPGNMKTSC-----FNVGNS-KCDGMNGWEVAGDHI- 473

```
Db 1577 CNSCHKG-----YYK-----DQKQVCKLCPNHFVCGKSKQKNDKNNISQYAGDAII 1622
QY 474 -----QSCAGGSDMDYLTLNLH-----IPG 493
Db 1623 CPNYSVTLOPENNISFKNCLIKGYENFQDFYNINNHCKAPLNFYKDTISNDLSP- 1681
QY 494 FKPTTSMTGATGSELGRITTFVETLCSADCVLVFMVDINRKTSTNVVSGGTKEKQAYTH 553
Db 1682 -CPTNSITLNTGA-----TSIYNCICDKG---FFYDVTYSVCNCPHG-YCSEKDMTTK 1731
QY 554 I-----IFKNATFTTFAFORTNOQDNRFRINDVWKIYSITATNAVGVASSCRACALG 608
Db 1732 LAPPIKCPKNYITTKGYSINTSHCVESGYTNTVIVEHS--SENGIIESHESAKNKIKLQ 1790
QY 609 SROSGSSVCPGPGHYIKETYN-QCKECPDPT-----YLSIHQVYKEACI 653
Db 1791 AKNRTSVCKDQSSYKSKISNEQCHKCPKSKTKLDFNNSDIFFCCLTMGYTDTKKECK 1850
QY 654 PCG-----PGSKNNQDHSVCYSDCFFYHEKENQILHYDFSNLVSVGLMNGSPSTSGT 707
Db 1851 PCWFENKLYCEGEKIYQIETIYDEILNIQYIRLLSDISTKKSFTIIN--EYINKGL 1908
QY 708 KYFHFENTSLCHGEKKKALCTNNITDTVKEIVAGSDDYT-----NLVCAFV-----CQS 758
Db 1909 -ISEITHISMMEKSKRKKSG---NNIQ--SIRKAKKDSEIYAVNIYRNLGNIITKORELKT 1962
QY 759 TIIPSEKGFRAALSQSITLADTFIGVTVEVTL---KNINIKEDMFVPVPTSQIPDVHFF 815
Db 1963 PKINNOIKKIELAKSIENKLIAN---INSKLTLLSEGNKNDTD-----TSIL--ANLF 2011
QY 816 YKSS-----TATTSCIN-GRSTAVKMRNPTKSGAGVISVPS-----KCPAGTCDGCTFYF 865
Db 2012 YKSSENIYIKHOKLINCORNVIPL-----GVDSSQNFDDCKCKKG-----YY 2055
QY 866 L-----WESAEACPLCTEHDHFEIG-----ACKRGFOETL 896
Db 2056 LEDRVLNKKIKCKPCPEGTFFNYGDVVKICSCPPKSTSIKGSYIPNHCFCRNGF---F 2112
QY 897 YVWNEPKWCIKIGISLPEKKLATC 919
Db 2113 YSKDTCLELEG-----ATC 2127

RESULT 12
Q99ND0 PRELIMINARY; PRT: 5374 AA.
ID Q99ND0;
AC Q99ND0;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE ZAN (Zonadhesin).
GN ZAN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RC MEDLINE=21138439; PubMed=11239002;
RA Wilson M.D., Klemmer C., Martindale D.W., Schnupf P., Boright A.P.,
RA Cheung T.L., Hardy D.M., Schwartz S., Scherer S.W., Tsui L.-C.,
RA Miller W., Koop B.F.;
RT "Comparative analysis of the gene-dense ACHIE/TFR2 region on human
RT chromosome 7q22 with the orthologous region on mouse chromosome 5.";
RL Nucleic Acids Res. 29:1352-1365(2001).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RA Cheung T.L., Wilson M.D., Koop B.F., Hardy D.M.;
RT "Genomic Basis of Inter- and Intra-species Variation in Zonadhesin
RT Domain Structure.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
```

```
CC -!- SIMILARITY: CONTAINS 3 MAM DOMAINS.
DR EMBL; AF312033; AAK28824.1; -.
DR EMBL; AY046056; AAL04416.1; -.
DR MGD; MGI:106656; ZAN.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR003645; FOLN.
DR InterPro; IPR000998; MAM_Domain.
DR InterPro; IPR003328; TILA_Cysrich.
DR InterPro; IPR002919; TIL_Cysrich.
DR InterPro; IPR001007; VWF-C.
DR InterPro; IPR001846; VWF-D.
DR Pfam; PF00629; MAM; 3.
DR Pfam; PF01826; TIL; 25.
DR Pfam; PF02345; TILA; 25.
DR Pfam; PF00094; vwd; 4.
DR SMART; SM00001; EGF_Like; 1.
DR SMART; SM00274; FOLN; 21.
DR SMART; SM00137; MAM; 3.
DR SMART; SM00214; VWC; 25.
DR SMART; SM00216; VWD; 4.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 18.
DR PROSITE; PS50060; MAM_2; 3.
KW EGF-like domain; Glycoprotein.
SQ SEQUENCE 5374 AA; 902D8CFE5DE24EB CRC64;

Query Match 3.6%; Score 193.5; DB 11; Length 5374;
Best Local Similarity 19.3%; Pred. No. 2.3e-06;
Matches 233; Conservative 103; Mismatches 387; Indels 487; Gaps 64;

QY 23 RGRSPWSPAWI-----CCWALAGCQ-----AAWAGDLPSSSR---PLPPQOE 63
Db 3122 OGALIEPKTWTISCTOSCNCTGGAIOCONFQCPKLYCKDLKDGSSNCTNIPL-----Q 3177
QY 64 KDYHPEYECDSGSRWRVAIPNSAVDCSGL---PDVVRGKECTFSC-ASGEYLEMKNO- 118
Db 3178 CPAHSRYTNCLPS-----CPPSLDPGLCEGTSPKVPSTCREGICQCPYLMHKNKC 3230
QY 119 -----VCS-----KCGEYISLGSGIKFDE 138
Db 3231 VLRIFCGCKNTGAFISADKTIWSRGCTQSCITCSAGAIHCRNFKCPGTSYKNGDNGSSN 3290
QY 139 WDELPAFNSNIATFMDTVVGPDSRDPDGNSSWTPRGNYIESNRDDCTVSLIYAVHLKK 198
Db 3291 CTEITLQCPNTSOFDCL-----PSCVPSCSNRCEVTSPSPSSCREGC-----LCN 3337
QY 199 SGYVFVEYQYVDNFIFFEFFIQNDQCOEMDTTDXVKVLTNDGEMGSHSVMLKSGTNILY 258
Db 3338 HGFVFSE-----DKCVPRTQCGCKDARGAIIAG----- 3366
QY 259 WRTTGILMGSKAVKPVLVKNITIEGVAYTSECPCKPGTFSNKPQSFNCQ--VCPRNTYS 316
Db 3367 -----KTWTSKGCTQSCACV-----EGNIQCONFCQCPPEYV- 3397
QY 317 EKGAKECIRCKDDSQFSGSSECTE---RPPCTTKDYFQIHTPC-----DEEGKTOI 364
Db 3398 -----CKDNSE--GSSTCTKITLQCPAHTQ-----YTSLPCLPCLPCLDPEG----- 3437
QY 365 MYKWIPIKI---CRE-----DLTDAI--RLPPSGEKDKDPPCPNPG 399
Db 3438 LCKDISPKVPSTCKEKGVCQSGYVLNSDKCVLRAECDCKDAOGALIPACKTWTSPGCTQS 3497
QY 400 -FYNNSSSSCH--PCPPGTF---SDGTKECR-----PCPA----- 428
Db 3498 CACMGAVQCQSSQCCPPGTYCKDNEIDGNSCAKITLQCPAHSFLTNCPLSCLPCLDPPG 3557
QY 429 ---GTEPALGFYKWNVLPNGNKTSC-----FNVNSKSCDGMN----- 464
Db 3558 LCKGASPK-----VPSCKECCICQSGYVLNSNCKLLRNRCCKDAHALIPEDK 3607
QY 465 -----GWEEV-----AGDHIOSSGAGGSNDYLI-----LNLHI----- 491
Db 3608 TWVSRGCTQSCVCTGGSIQCLSPQCPGPGAYCKDNEIDGNSNCARIPQCPANSHYTDGFP 3667
```



```
DR EMBL; Z81125; CAAL5432.2; JOINED.
DR EMBL; Z81125; CAB03385.2; -.
DR EMBL; AL008585; CAB03385.2; JOINED.
DR HSSP; P02468; 1TLE.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000034; Laminin_B.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR001886; LamNT.
DR Pfam; PF00052; laminin_B; 2.
DR Pfam; PF00053; laminin_EGF; 17.
DR Pfam; PF00054; laminin_G; 1.
DR Pfam; PF00055; laminin_Nterm; 1.
DR PRINTS; PR00011; EGF_LAMININ.
DR ProDom; PD002082; LamNT; 1.
DR ProDom; PD003031; Laminin_B; 2.
DR SMART; SM00180; EGF_Lam; 14.
DR SMART; SM00281; LamB; 2.
DR SMART; SM00282; LamG; 2.
DR SMART; SM00136; LamNT; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_11.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 14.
FT NON_TER 2759
SQ SEQUENCE 2759 AA; 305515 MW; 081F737000C63541 CRC64;

Query Match 3.6%; Score 192.5; DB 5; Length 2759;
Best Local Similarity 18.8%; Pred. No. 1.1e-06;
Matches 205; Conservative 96; Mismatches 326; Indels 461; Gaps 59;

QY 39 LAGCAQAGADLPSSSRPLPPC-----QEKDY-----HFEYTECDSSGS 78
DB 819 LGACEQC---ECPSLDLPNPEICISTELAVLGVSASNEEDNYVCINPLGYEGNKCEY--- 872

QY 79 RWRVAIPNSAVDCSG--LPDPVRGK--ECTF-----SCASGEYLEMKNOVCSKC- 123
DB 873 -----CSGFTEDPLTGICIECTCNGNDTPMGIGNCDS-----ETGKCLKCI 914

QY 124 GEGTYSLGSIGKIFDEWDELPAQFSNIATFMDTVVGPSSRSRDPDCNNSSWIPRGNVIESNR 183
DB 915 GHTTGDCSCESEKHHW-----GNAQLT-----CKPGCGHTQGAENVQCSENGE 959

QY 184 DCTVSLIYAVHLKSGYVFYQVVDNIFPEFIQNDQCEQMDTTDDKWKVLTNDNGEW 243
DB 960 CECKENYIGA-----QCDRCENHGDVENGGPCACDCNDT 993

QY 244 GSHSVMLKSGTNLYWRTTGILMGSKAVKPVLVKNITIEGVAYTSECFCKPGTFESN--- 300
DB 994 GS-----IGSDCDQ-----VSGQC-NCKOGVFGKQCD 1019

QY 301 --KPGSFN-----CQVCPRNTYSEKAKERICRDKDSDQSGSSECTERPPCT-----TKD 348
DB 1020 QCRPSYFNFTDAGCFCHNIY--GSIEDGKC---DQTGKCECENVEGTMCEKCADG 1073

QY 349 YFOIHT-----PDEEGKTQIMYKWIPEK-ICREDLTDAIRLPPSGKKDCCPCPNPGF 400
DB 1074 YFNITSGDCEGCDGPTGSEDVSNLVTGQCVCQKPGVT-----GLK--CDSCLPNF 1123

QY 401 Y---NNGSSSCHPCP--PGTFSFGTKECRCPAGTEPALGFEYKWNVLPGNMKTSCFNVG 456
DB 1124 YGLTSEGCTECPGAPGQVCPIDGSCVCPNPT-----VGEMCENCTT--- 1167

QY 457 NSKCDGMNWEVAGDHIOGAGSGSDNDYLILNLHPIPGFKPPTSMTGATGSELGRITFVPE 516
DB 1168 -----NAW-----DYHPLN-----GCK----- 1179

QY 517 TLCSADCVLYFMVDINRKSTNVVSWGGTKEKQAY-----THIFKNATFTTWARQ 568
DB 1180 -LC--DC-----SDIGSDGGMGNTFTGQCKCKAAVYGLKCDLCTHGFNFPT----- 1223

QY 569 RTNOGDNRREINDMVKIYSITATNAVDGVASSCRACALGSEQSGSSVCPPGPHYIEKE 628
DB 1224 -----CEPC-----GCNAAGTDPLOCKDGQCL--- 1245
```

```
QY 629 TNQCKECPDPTVLSIHQVYGKEACIPCGPG-----SKNQDHSVCYSDCF----- 673
DB 1246 CNEIGECP-----CKKNVHGK-CDQCCEGTFSLDSSNLKGTCECF--CFNRTSNCEQSD 1297

QY 674 -----FYHEKEN-QILHYDFSNLSS-----VGSILMNGPS 701
DB 1298 LVWQOYAEADRRRAVQEPWEFYTKKHINILLREKPSHFNSYPTDATPLYWPLPSTMLGDR 1357

QY 702 FTS-----KGTKEY-HFENISLCGHEGKKMALCTNNITDFTVK 738
DB 1358 TASYNGFLREFKIWNEDNRRLGHLGIRPDQOYFRHFOVFIQGNRIELEHPIHMEINDDGIY 1417

QY 739 EI-----VAGSDDYT-----NLVCAFPVCOSTIIPSESKGFRAALSSQSIL 779
DB 1418 KIRLHESEWRVRHSPELTLTRKQMMVALQDTQGIYIRGTYYTPARGDAINIQVSLDVAV 1477

QY 780 ADTFIGVIVETTLKNINIKEDM-----FPVPYISQIPDVHFFYKSS----- 819
DB 1478 PESKIVAGLSTT-KAIGVEKCLGCPGYTGLSCQPEVGYRKKHREYLNQADIALIGW 1536

QY 820 TATTSCINGRSTAVKMRNPTKSGAGVISVPSKCPAGT-----CDGCTFYFLWESAE---- 871
DB 1537 SEPCSC-HGHS-----QTCNPD-----TSVCTDCHEINTFGDCEHCLPGYIGDAREGGAN 1585

QY 872 -----ACPL-----CTEHPFHEIEGACKRGFQETLYVMNPKWCIKG-ISLPEKK 915
DB 1586 ACTKACACPLVENSFSFSDSCVAVDHGRGYVCSCKPG-----YTGQYCETCVAGYGDPOHI 1640

QY 916 LATCETVD 923
DB 1641 GGTCSPCD 1648

RESULT 15
QYTR4 ID Q9TR4 PRELIMINARY; PRT: 3102 AA.
AC Q9TR4;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update).
DE Laminin alpha chain.
GN LAM1/2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Zhu X., Kao G., Joh K., Hall D.H., Wadsworth V.K., Hutter H.,
RA Vogel B.E., Huang C., Yochem J., Horl K., Hedgecock E.M.,
RA Wadsworth W.G.;
RT "Expression, function and evolution of laminin alpha chains.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR HSSP; P02468; 1TLE.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000034; Laminin_B.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR001886; LamNT.
DR Pfam; PF00052; laminin_B; 2.
DR Pfam; PF00053; laminin_EGF; 17.
DR Pfam; PF00054; laminin_G; 3.
DR Pfam; PF00055; laminin_Nterm; 1.
DR PRINTS; PR00011; EGF_LAMININ.
DR ProDom; PD002082; LamNT; 1.
DR ProDom; PD003031; Laminin_B; 2.
DR SMART; SM00180; EGF_Lam; 14.
DR SMART; SM00001; EGF_like; 2.
DR SMART; SM00281; LamB; 2.
DR SMART; SM00282; LamG; 4.
```


GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 13:31:24 ; Search time 17.5467 Seconds
(without alignments)
5626.691 Million cell updates/sec

Title: US-10-073-333A-4

Perfect score: 1027

Sequence: 1 MLFRARGVVRGWRGPAEA.....KEKEDHFESVQKTSRSPNI 1027

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_73:*

1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	0.8	100	1 S27016	defensin alpha-6 p
2	8	0.8	148	2 E84239	hypothetical prote
3	8	0.8	184	1 QOCVLG	BLI protein - toma
4	8	0.8	228	2 S65426	pyruvate decarboxy
5	8	0.8	228	2 S65425	pyruvate decarboxy
6	8	0.8	239	2 G87265	conserved hypothet
7	8	0.8	268	2 A81424	E. coli RpiR trans
8	8	0.8	268	2 A81797	E. coli RpiR trans
9	8	0.8	271	2 S12783	OX40 antigen precu
10	8	0.8	272	2 I48700	gene ox40 protein
11	8	0.8	281	2 T33466	hypothetical prote
12	8	0.8	286	2 A82929	ATP synthase gamma
13	8	0.8	405	2 S65471	pyruvate decarboxy
14	8	0.8	417	2 S57820	pyruvate decarboxy
15	8	0.8	428	2 A70241	probable coenzyme
16	8	0.8	444	2 T01548	protein kinase hom
17	8	0.8	463	2 T28748	hypothetical prote
18	8	0.8	585	2 T03252	pyruvate decarboxy
19	8	0.8	603	2 T03295	pyruvate decarboxy
20	8	0.8	622	2 A36915	fructanase - Bacte
21	8	0.8	671	2 T10755	kinesin-related pr
22	8	0.8	742	2 A82357	hypothetical prote
23	8	0.8	881	2 T01269	serine/threonine-s
24	8	0.8	1206	2 B87247	probable conserved
25	8	0.8	1578	2 A81512	peptidoglycan boun
26	8	0.8	1582	2 AC1153	adhesin homolog lm
27	8	0.8	3102	2 T43291	laminin alpha chal
28	7	0.7	94	2 A64863	hypothetical prote
29	7	0.7	100	1 B43719	urease (EC 3.5.1.5

30: 7 0.7 100 1 S08478 urease (EC 3.5.1.5)
31 7 0.7 100 1 TNLJSI trans-activating t
32 7 0.7 100 2 B47090 urease (EC 3.5.1.5)
33 7 0.7 100 2 A43998 hypothetical prote
34 7 0.7 109 2 S42599 hypothetical prote
35 7 0.7 122 2 T28199 hypothetical prote
36 7 0.7 134 2 C84023 hypothetical prote
37 7 0.7 134 2 T36365 proline-rich prote
38 7 0.7 135 2 PC2269 cytochrome P450 pr
39 7 0.7 141 2 A27482 vasotocin / neurop
40 7 0.7 147 2 E70746 probable mmp52 pro
41 7 0.7 157 2 AG3560 transcription regu
42 7 0.7 159 2 B29879 vasotocin / neurop
43 7 0.7 173 2 B3450 invasion protein b
44 7 0.7 179 2 H64472 hypothetical prote
45 7 0.7 180 2 T49530 related to glycine

ALIGNMENTS

RESULT 1
S27016
defensin alpha-6 precursor - human
N:Alternate names: Paneth cell-specific alpha-defensin 6
C:Species: Homo sapiens (man)
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: S27016
R:Jones, D.E.; Bevins, C.L.
FEBS Lett. 315, 187-192, 1993
A:Title: Defensin-6 mRNA in human Paneth cells: Implications for antimicrobial peptid
A:Reference number: S27016; MUID:93114459; PMID:841797
A:Accession: S27016
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-100 <JON>
A:Cross-references: GB:M98331; NID:g181546; PIDN:AAB59357.1; PID:g181547
C:Genetics:
A:Gene: GDB:DEFA6; DEF6; HD-6
A:Cross-references: GDB:136838; OMIM:600471
A:Map position: 8pter-8p21
C:Superfamily: mammalian defensin
C:Keywords: antibiotic; homodimer
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-70/Domain: propeptide #status predicted <PRO>
F:71-100/product: defensin alpha-6 #status predicted <MAT>
F:72-99,74-88,78-98/Disulfide bonds: #status predicted
Query Match 0.8%; Score 8; DB 1; Length 100;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 936 TAVLLVAL 943
DB 8 TAVLLVAL 15
RESULT 2
E84239
hypothetical protein Vng0825c [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: E84239
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky
; Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.;
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: E84239
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-148 <STO>
 A:Cross-references: GB:AF004437; NID:gl0580395; PIDN:AAG19281.1; GSPDB:GN00138
 C:Genetics:
 A:Gene: VNG0825C

Query Match 0.8%; Score 8; DB 2; Length 148;
 Best Local Similarity 100.0%; Pred. No. 6.0;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 155 TVWGPSDS 162
 |||||
 Db 5 TVWGPSDS 12

RESULT 3

QOCVLG

BL1 protein - tomato golden mosaic virus

C:Species: tomato golden mosaic virus

A:Note: host Nicotiana sp. (tobacco)

C:Date: 28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change 08-Apr-1994

C:Accession: A04169

R:Hamilton, W.D.O.; Stein, V.E.; Coutts, R.H.A.; Buck, K.W.

EMBO J. 3, 2197-2205, 1984

A:Title: Complete nucleotide sequence of the infectious cloned DNA components of tomato

A:Reference number: A04163

A:Accession: A04169

A:Molecule type: DNA

A:Residues: 1-184 <HAM>

C:Comment: The genome consists of two circular, single-stranded DNA components, DNA A and

C:Genetics:

A:Map position: segment B

C:Superfamily: tomato golden mosaic virus BL1 protein

Query Match 0.8%; Score 8; DB 1; Length 184;

Best Local Similarity 100.0%; Pred. No. 7.9;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 177 NVIESNRD 184
 |||||
 Db 13 NVIESNRD 20

RESULT 4

S65426

pyruvate decarboxylase (EC 4.1.1.1) - fava bean (fragment)

C:Species: Vicia faba (fava bean)

C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 15-Oct-1999

C:Accession: S65426

R:Muecke, U.; Wohlfarth, T.; Fiedler, U.; Baumelein, H.; Ruecknagel, K.P.; Koenig, S.

Eur. J. Biochem. 237, 373-382, 1996

A:Title: Pyruvate decarboxylase from Pisum sativum. Properties, nucleotide and amino acid

A:Reference number: S65423; MUID:96215432; PMID:8647075

A:Accession: S65426

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-228 <MU>

C:Superfamily: acetolactate synthase large chain; thiamin pyrophosphate-binding domain

C:Keywords: carbon-carbon lyase; carboxy-lyase

Query Match 0.8%; Score 8; DB 2; Length 228;

Best Local Similarity 100.0%; Pred. No. 9.4;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 269 KAVKPVLV 276
 |||||
 Db 50 KAVKPVLV 57

RESULT 5

S65425

pyruvate decarboxylase (EC 4.1.1.1) (clone PDC3) - garden pea (fragment)

C:Species: Pisum sativum (garden pea)

C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 15-Oct-1999

C:Accession: S65425

R:Muecke, U.; Wohlfarth, T.; Fiedler, U.; Baumelein, H.; Ruecknagel, K.P.; Koenig, S.

Eur. J. Biochem. 237, 373-382, 1996

A:Title: Pyruvate decarboxylase from Pisum sativum. Properties, nucleotide and amino

A:Reference number: S65423; MUID:96215432; PMID:8647075

A:Accession: S65425

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-228 <MU>

C:Superfamily: acetolactate synthase large chain; thiamin pyrophosphate-binding domain

C:Keywords: carbon-carbon lyase; carboxy-lyase

Query Match 0.8%; Score 8; DB 2; Length 228;

Best Local Similarity 100.0%; Pred. No. 9.4;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 269 KAVKPVLV 276
 |||||
 Db 50 KAVKPVLV 57

RESULT 6

G87265

conserved hypothetical protein CC0136 [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C:Accession: G87265

R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,

B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko

n, J.; Emolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: G87265

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-239 <STO>

A:Cross-references: GB:AF005673; NID:gl3421247; PIDN:AAK22123.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC0136

Query Match 0.8%; Score 8; DB 2; Length 239;

Best Local Similarity 100.0%; Pred. No. 9.8;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 AEAPRRGR 25
 |||||
 Db 176 AEAPRRGR 183

RESULT 7

AB1424

E. coli RpiR transcription regulator homolog lmo2795 [imported] - Listeria monocytoge

C:Species: Listeria monocytogenes

C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001

C:Accession: AB1424

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec

; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsthi,

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;

Ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boiland, J.A.; Voss, H.; Wehla

A:Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AB1424

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-268 <GLA>

A:Cross-references: GB:NC_003210; PIDN:CAD01008.1; PID:gl6412295; GSPDB:GN00177

A:Experimental source: strain EGD-e

C:Genetics:

A:Gene: lmo2795

Query Match 0.8%; Score 8; DB 2; Length 268;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 792 LKNINKE 799
 |||||
 DB 32 LKNINKE 39

RESULT 8
 AH1797
 E. coli RpiR transcription regulator homolog lin2927 [imported] - Listeria innocua (strain
 C:Species: Listeria innocua
 C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
 C:Accession: AH1797
 R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
 D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
 Science 294, 849-852, 2001
 A:Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
 ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
 A:Title: Comparative genomics of Listeria species.
 A:Reference number: AB1077; MUID:21537279; PMID:11679669
 A:Accession: AH1797
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-268 <GLA>
 A:Cross-references: GB:AL592022; PIDN:CAC98152.1; PID:gl6415468; GSPDB:GN00178
 A:Experimental source: strain Clp11262
 C:Genetics:
 A:Gene: lin2927

Query Match 0.8%; Score 8; DB 2; Length 268;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 792 LKNINKE 799
 |||||
 DB 32 LKNINKE 39

RESULT 9
 SI2783
 OX40 antigen precursor - rat
 N:Alternate names: nerve growth factor receptor homolog
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 05-Nov-1999
 C:Accession: SI2783; S08036
 R:Mallett, S.; Fossum, S.; Barclay, A.N.
 EMBO J. 9, 1063-1068, 1990
 A:Title: Characterization of the MRC OX40 antigen of activated CD4 positive T lymphocyte
 A:Reference number: SI2783; MUID:90214614; PMID:2157591
 A:Accession: SI2783
 A:Molecule type: mRNA
 A:Residues: 1-271 <MAL>
 A:Cross-references: EMBL:X17037; NID:G57830; PIDN:CAA34897.1; PID:G57831
 C:Superfamily: CD27 antigen; NGF receptor repeat homology
 C:Keywords: growth factor receptor; transmembrane protein
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-271/Product: OX40 antigen #status predicted <MAT>
 F:211-235/Domain: transmembrane #status predicted <TMM>

Query Match 0.8%; Score 8; DB 2; Length 271;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCPGGH 623
 |||||
 DB 122 CVPCPGGH 129

RESULT 10

I48700
 gene ox40 protein - mouse
 N:Alternate names: OX40 antigen
 C:Species: Mus musculus (house mouse)
 C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 11-Jan-2000
 C:Accession: I48700; I48334; S34377
 R:Calderhead, D.M.; Buhlmann, J.E.; van den Eertwegh, A.J.; Claassen, E.; Noelle, R.J
 J. Immunol. 151, 5261-5271, 1993
 A:Title: Cloning of mouse OX40: a T cell activation marker that may mediate T-B cell
 A:Reference number: I48700; MUID:94044750; PMID:8228223
 A:Accession: I48700
 A:Status: translated from GB/EMBL/DBDJ
 A:Molecule type: mRNA
 A:Residues: 1-272 <RES>
 A:Cross-references: EMBL:Z21674; NID:G312827; PIDN:CAA79772.1; PID:G312828
 R:Birkeland, M.B.; Copeland, N.G.; Gilbert, D.J.; Jenkins, N.A.; Barclay, A.N.
 Eur. J. Immunol. 25, 926-930, 1995
 A:Title: Gene structure and chromosomal localization of the mouse homologue of rat OX
 A:Reference number: I48334; MUID:95255413; PMID:7737295
 A:Accession: I48334
 A:Status: translated from GB/EMBL/DBDJ
 A:Molecule type: DNA
 A:Residues: 1-14, 'G', 16-272 <RE2>
 A:Cross-references: EMBL:X85214; NID:G732818; PIDN:CAA59476.1; PID:G732819
 C:Genetics:
 A:Gene: ox40
 A:Introns: 45/1; 86/1; 122/1; 144/2; 210/1; 250/1
 C:Superfamily: CD27 antigen; NGF receptor repeat homology

Query Match 0.8%; Score 8; DB 2; Length 272;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCPGGH 623
 |||||
 DB 123 CVPCPGGH 130

RESULT 11
 T33466
 hypothetical protein C43H8.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T33466
 R:Tin-Wollam, A.M.
 submitted to the EMBL Data Library, October 1998
 A:Description: The sequence of C. elegans cosmid C43H8.
 A:Reference number: 221350
 A:Accession: T33466
 A:Status: preliminary; translated from GB/EMBL/DBDJ
 A:Molecule type: DNA
 A:Residues: 1-281 <TIN>
 A:Cross-references: EMBL:AF098499; PIDN:AAC67398.1; GSPDB:GN000019; CESP:C43H8.3
 A:Experimental source: strain Bristol N2; clone C43H8
 C:Genetics:
 A:Gene: CESP:C43H8.3
 A:Map position: 1
 A:Introns: 58/1; 114/2; 242/3

Query Match 0.8%; Score 8; DB 2; Length 281;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 731 NITDFTVK 738
 |||||
 DB 111 NITDFTVK 118

RESULT 12
 AB2929
 ATP synthase gamma chain UUI30 [imported] - Ureaplasma urealyticum
 C:Species: Ureaplasma urealyticum
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000

C:Accession: A82929
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.
Submitted to GenBank, February 2000
A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a min
A:Reference number: A82870
A:Accession: A82929
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-286 <GLA>
A:Cross-references: GB:AE002114; GB:AF222894; NID:g6899086; PIDN:AAF30536.1; GSPDB:GN001
A:Experimental source: serovar 3; biovar 1
C:Genetics:
A:Gene: atpG; UUI30
A:Genetic code: SOC3
C:Superfamily: H+-transporting ATP synthase gamma chain

Query Match 0.8%; Score 8; DB 2; Length 286;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 739 EIVAGSD 746
|||||||
Db 279 EIVAGSD 286

RESULT 13
S65471
pyruvate decarboxylase (EC 4.1.1.1) (clone PDC2) - garden pea (fragment)
C:Species: Pisum sativum (garden pea)
C:Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 20-Jun-2000
C:Accession: S65471; S65424
R:Muecke, U.; Wohlfarth, T.; Fiedler, U.; Baeumlein, H.; Ruecknagel, K.P.; Koenig, S.
Eur. J. Biochem. 237, 373-382, 1996
A:Title: Pyruvate decarboxylase from Pisum sativum. Properties, nucleotide and amino acid
A:Reference number: S65423; MUID:96215432; PMID:8647075
A:Accession: S65471
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-405 <MU>
A:Cross-references: EMBL:Z66544; NID:g1177604; PIDN:CAA91445.1; PID:g1177605
A:Accession: S65424
A:Molecule type: protein
A:Residues: 1-8;146-153;209-218;350-362;386-393 <MU>
C:Genetics:
A:Gene: pdc
C:Superfamily: acetolactate synthase large chain; thiamin pyrophosphate-binding domain
C:Keywords: carbon-carbon lyase; carboxy-lyase
F:1-405/Product: pyruvate decarboxylase #status experimental <MAT>
F:271-317/Domains: thiamin pyrophosphate-binding domain homology <TPB>

Query Match 0.8%; Score 8; DB 2; Length 405;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 KAVKPVLV 276
|||||||
Db 50 KAVKPVLV 57

RESULT 14
S57820
pyruvate decarboxylase (EC 4.1.1.1) 1 - common tobacco (fragment)
C:Species: Nicotiana tabacum (common tobacco)
C:Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 15-Oct-1999
C:Accession: S57820
R:Bucher, M.; Brander, K.A.; Sbicego, S.; Mandel, T.; Kuhlmeier, C.
Plant Mol. Biol. 28, 739-750, 1995
A:Title: Aerobic fermentation in tobacco pollen.
A:Reference number: S57819; MUID:95375236; PMID:7647304
A:Accession: S57820
A>Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-417 <BUC>

A:Cross-references: EMBL:X81854
C:Superfamily: acetolactate synthase large chain; thiamin pyrophosphate-binding domain
C:Keywords: carbon-carbon lyase; carboxy-lyase

Query Match 0.8%; Score 8; DB 2; Length 417;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 KAVKPVLV 276
|||||||
Db 155 KAVKPVLV 162

RESULT 15
AF0241
probable coenzyme synthetase YPO1981 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C:Accession: AF0241
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.
ll, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barral
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AF0241
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-428 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC90794.1; PID:g15979994; GSPDB:GN00175
C:Genetics:
A:Gene: YPO1981

Query Match 0.8%; Score 8; DB 2; Length 428;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 604 ACALGSEQ 611
|||||||
Db 84 ACALGSEQ 91

Search completed: May 12, 2003, 13:41:40
Job time : 19.5467 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 13:23:58 ; Search time 13.9342 Seconds
(without alignments)
3056.959 Million cell updates/sec

Title: US-10-073-333A-4

Perfect score: 1027

Sequence: 1 MLFRARGPVRGGRPAEA.....KEKEDHFESVQLKTSRSPNI 1027

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8	0.8	100	1 DEF6_HUMAN	Q01524 homo sapien
2	8	0.8	271	1 TNR4_RAT	P15725 ratus norv
3	8	0.8	272	1 TNR4_MOUSE	P47741 mus musculus
4	8	0.8	293	1 VBL1_TCMV	P03566 tomato gold
5	8	0.8	300	1 TR6B_HUMAN	O95407 homo sapien
6	8	0.8	405	1 DCP2_PEA	P51851 pisum sativ
7	8	0.8	418	1 DCP1_TOBAC	P51845 nicotiana t
8	8	0.8	430	1 TRUT_HUMAN	Q96924 homo sapien
9	8	0.8	430	1 TRUT_MACFA	O9n092 macaca fasc
10	8	0.8	585	1 DCP3_ORYSA	P51849 oryza sativ
11	8	0.8	603	1 DCP2_ORYSA	P51848 oryza sativ
12	8	0.8	671	1 KRP2_RAT	Q62909 ratus norv
13	8	0.8	718	1 MCAK_CRIGR	P70096 cricetus
14	7	0.7	100	1 TAT_SIVCZ	P17285 chimpanzee
15	7	0.7	100	1 URE3_ECOLI	Q03282 escherichia
16	7	0.7	100	1 URE3_PROMI	P17088 proteus mir
17	7	0.7	100	1 URE3_PROVU	P16124 proteus vul
18	7	0.7	101	1 RK24_GUITH	O46905 guillardia
19	7	0.7	134	1 CRB2_BACHD	Q9k819 bacillus ha
20	7	0.7	141	1 NEUV_RANES	P11858 rana esculi
21	7	0.7	145	1 MMS2_MYCTU	Q11170 mycobacteri
22	7	0.7	159	1 NEUV_BUFJA	P08163 bufo japoni
23	7	0.7	179	1 YCBU_ECOLI	P75855 escherichia
24	7	0.7	179	1 YD85_METJA	Q58780 methanococ
25	7	0.7	200	1 RS8A_SCHPO	O14049 schizosacch
26	7	0.7	200	1 RS8B_SCHPO	Q9p7b2 schizosacch
27	7	0.7	203	1 I296_METJA	Q57744 methanococ
28	7	0.7	206	1 Y930_HAEIN	P44077 haemophilus
29	7	0.7	208	1 YMI6_YEAST	P40206 saccharomyc
30	7	0.7	208	1 YQED_BACSU	P54449 bacillus su
31	7	0.7	259	1 SP07_YEAST	P18410 saccharomyc
32	7	0.7	270	1 PYRF_CANAL	P13649 candida alb
33	7	0.7	270	1 PYRF_CANDU	Q9c150 candida dub

RESULT 1

ID	DEF6_HUMAN	STANDARD;	PRT;	100 AA.
AC	Q01524;			
DT	01-JUL-1993 (Rel. 26, Created)			
DT	01-JUL-1993 (Rel. 26, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Defensin 6 precursor (Defensin, alpha 6).			
GN	DEFA6 OR DEF6.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Intestine;			
RX	MEDLINE=93114459; PubMed=8417977;			
RA	Jones D.E., Bevins C.L.;			
RT	"Defensin-6 mRNA in human Paneth cells: implications for antimicrobial peptides in host defense of the human bowel.";			
RL	FEBS Lett. 315:187-192(1993).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=96223969; PubMed=8626737;			
RA	Mallow E.B., Harris A., Salzman N., Russell J.P.,			
RA	Deberardinis R.J., Richelli E., Bevins C.L.;			
RT	"Human enteric defensins. Gene structure and developmental expression.";			
RL	J. Biol. Chem. 271:4038-4045(1996).			
CC	-!- FUNCTION: HAS ANTIMICROBIAL ACTIVITY.			
CC	-!- TISSUE SPECIFICITY: PANETH CELLS OF THE SMALL INTESTINE.			
CC	-!- SIMILARITY: BELONGS TO THE CORTICOSTATIN/DEFENSIN FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; M98331; AAB59357.1;			
DR	EMBL; U33317; AAC50382.1; ALT_SEQ.			
DR	PIR; S27016; S27016.			
DR	Genew; HGNC:2765; DEFA6.			
DR	MIM; 600471;			
DR	InterPro; IPR002366; Defensin.alpha.			
DR	InterPro; IPR001271; Defensin_mammal.			
DR	Pfam; PF00323; defensins; 1.			
DR	Pfam; PF00879; Defensin_propep; 1.			
DR	SMART; SM00048; DEFSN; 1.			
DR	PROSITE; PS00269; DEFENSIN; 1.			
KW	Defensin; Antibiotic; Fungicide; Signal.			
FT	SIGNAL 1 19 POTENTIAL.			
FT	PROPEP 20 265 POTENTIAL.			
FT	CHAIN 266 100 DEFENSIN 6.			

34	7	0.7	277	1	ATND_XENLA	P21188 xenopus lae
35	7	0.7	283	1	TRI4_HUMAN	Q92956 homo sapien
36	7	0.7	320	1	DNC_HUMAN	Q9hc21 homo sapien
37	7	0.7	335	1	FLIG_THEMEA	Q9wy63 thermotoga
38	7	0.7	362	1	YD33_MYCLE	P53425 mycobacteri
39	7	0.7	365	1	WNI4_HUMAN	O14904 homo sapien
40	7	0.7	418	1	YEB7_SCHPO	O14303 schizosacch
41	7	0.7	420	1	YCEL_CAEEL	Q94175 caenorhabdi
42	7	0.7	437	1	YOEZ_BACSU	P54465 bacillus su
43	7	0.7	459	1	G33_RAT	P05432 rattus norv
44	7	0.7	463	1	YAD4_YEAST	P28003 saccharomyc
45	7	0.7	529	1	YPCI_CAEEL	Q11178 caenorhabdi

ALIGNMENTS

FT DISULFID 72 99 BY SIMILARITY.
 FT DISULFID 74 88 BY SIMILARITY.
 FT DISULFID 78 98 BY SIMILARITY.
 SQ SEQUENCE 100 AA; 10975 MW; EDF7E033DDCE2D5 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 100;
 Best Local Similarity 100.0%; Pred. No. 2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 936 TAVLLVAL 943

DB 8 TAVLLVAL 15

RESULT 2

TNR4_RAT ID TNR4_RAT STANDARD; PRT; 271 AA.
 AC P15725;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Tumor necrosis factor receptor superfamily member 4 precursor (OX40L receptor) (OX40 antigen) (MRC OX40).
 GN TNFRSF4 OR TXGPI1 OR OX40.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=F-cell;
 RX MEDLINE=90214614; PubMed=2157591;
 RA Mallett S., Fossum S., Barclay A.N.;
 RT "Characterization of the MRC OX40 antigen of activated CD4 positive T lymphocytes -- a molecule related to nerve growth factor receptor.";
 RL EMBO J. 9:1063-1068(1990).
 CC -!- FUNCTION: Receptor for TNFSF4/OX40L/GP34.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- TISSUE SPECIFICITY: ACTIVATED T-CELLS.
 CC -!- SIMILARITY: CONTAINS 4 TNFR-CYS REPEATS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch)

CC EMBL; X17037; CAA34897.1; -.
 CC PIR; S08036; S08036.
 CC PIR; S12783; S12783.
 CC HSP; O14763; 1D4V.
 CC InterPro: IPR001368; TNFR_c6.
 CC Pfam: PF00020; TNFR_c6; 3.
 CC ProDom: PD000771; TNFR_c6; 1.
 CC SMART; SM00208; TNFR; 3.
 CC PROSITE; PS00652; TNFR_NGFR_1; 2.
 CC PROSITE; PS00652; TNFR_NGFR_2; 2.
 CC Receptor; Antigen; Transmembrane; Glycoprotein; Repeat; Signal.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 271 TUMOR NECROSIS FACTOR RECEPTOR SUPERFAMILY MEMBER 4. EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 20 210 POTENTIAL.
 FT TRANSMEM 211 235 POTENTIAL.
 FT DOMAIN 236 271 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 25 60 TNFR-CYS 1.
 FT REPEAT 61 102 TNFR-CYS 2.
 FT REPEAT 103 123 TNFR-CYS 3 (INCOMPLETE).
 FT REPEAT 124 164 TNFR-CYS 4.
 FT DISULFID 26 37 BY SIMILARITY.
 FT DISULFID 38 51 BY SIMILARITY.
 FT DISULFID 41 59 BY SIMILARITY.

FT DISULFID 62 76 BY SIMILARITY.
 FT DISULFID 79 94 BY SIMILARITY.
 FT DISULFID 82 102 BY SIMILARITY.
 FT DISULFID 104 122 BY SIMILARITY.
 FT DISULFID 125 138 BY SIMILARITY.
 FT DISULFID 144 163 BY SIMILARITY.
 FT CARBOHYD 143 143 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 271 AA; 29895 MW; C06465136B16B821 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 271;
 Best Local Similarity 100.0%; Pred. No. 4.7;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCPEGH 623

DB 122 CVPCPEGH 129

RESULT 3

TNR4_MOUSE ID TNR4_MOUSE STANDARD; PRT; 272 AA.
 AC P47741;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Tumor necrosis factor receptor superfamily member 4 precursor (OX40L receptor) (OX40 antigen).
 GN TNFRSF4 OR TXGPI1 OR OX40.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/C.
 RX MEDLINE=94044750; PubMed=8228223;
 RA Calderhead D.M., Buhlmann J.E., van den Bortwegh A.J.,
 RA Claassen E., Noelle R.J., Fell H.;
 RT "Cloning of mouse OX40: a T cell activation marker that may mediate T-B cell interactions.";
 RL J. Immunol. 151:5261-5271(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95255413; PubMed=7737295;
 RA Birkeland M.L., Copeland N.G., Gilbert D.J., Jenkins N.A.,
 RA Barclay A.N.;
 RT "Gene structure and chromosomal localization of the mouse homologue of rat OX40 protein.";
 RL Eur. J. Immunol. 25:926-930(1995).
 CC -!- FUNCTION: Receptor for TNFSF4/OX40L/GP34.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- SIMILARITY: CONTAINS 4 TNFR-CYS REPEATS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch)

CC EMBL; Z21674; CAA79772.1; -.
 CC EMBL; X85214; CAA59476.1; -.
 CC HSP; O14763; 1D0G.
 CC MGD; MGI:104512; Tnfrsf4.
 CC InterPro: IPR001368; TNFR_c6.
 CC Pfam: PF00020; TNFR_c6; 3.
 CC ProDom: PD000771; TNFR_c6; 1.
 CC SMART; SM00208; TNFR; 3.
 CC PROSITE; PS00652; TNFR_NGFR_1; 2.
 CC PROSITE; PS00652; TNFR_NGFR_2; 2.
 CC Receptor; Antigen; Transmembrane; Glycoprotein; Repeat; Signal.
 FT SIGNAL 1 19 POTENTIAL.

```

FT CHAIN 20 272 TUMOR NECROSIS FACTOR RECEPTOR
FT DOMAIN 20 211 SUPERFAMILY MEMBER 4.
FT TRANSMEM 212 236 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 237 272 POTENTIAL.
FT REPEAT 26 61 CYTOPLASMIC (POTENTIAL).
FT REPEAT 62 103 TNFR-CYS 1.
FT REPEAT 104 124 TNFR-CYS 2.
FT REPEAT 125 165 TNFR-CYS 3. (INCOMPLETE).
FT REPEAT 166 208 TNFR-CYS 4.
FT DISULFID 27 38 BY SIMILARITY.
FT DISULFID 39 52 BY SIMILARITY.
FT DISULFID 42 60 BY SIMILARITY.
FT DISULFID 63 77 BY SIMILARITY.
FT DISULFID 80 95 BY SIMILARITY.
FT DISULFID 83 103 BY SIMILARITY.
FT DISULFID 105 123 BY SIMILARITY.
FT DISULFID 126 139 BY SIMILARITY.
FT DISULFID 145 164 BY SIMILARITY.
FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 15 15 A -> G (IN REF. 2).
SQ SEQUENCE 272 AA; 30153 MW; 06E7BB4156F0D08E CRC64;

Query Match 0.8%; Score 8; DB 1; Length 272;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 616 CVPCPGPGH 623
DB 123 CVPCPGPGH 130

RESULT 4
VBL1_TGMV
ID VBL1_TGMV STANDARD; PRT; 293 AA.
AC P03566;
DC 21-JUL-1986 (Rel. 01, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE BL1 protein.
GN BCL1.
OS Tomato golden mosaic virus (TGMV).
OC Viruses; ssDNA viruses; Geminiviridae; Begomovirus.
OX NCBI_TaxID=10831;
RN [1]
RP SEQUENCE FROM N.A.
RA von Arnim A.G., Stanley J.;
RL Submitted (XXX-1991) to the EMBL/GenBank/DBJ databases.
RN [2]
RP PRELIMINARY SEQUENCE FROM N.A.
RA Hamilton W.D.O., Stein V.E., Coutts R.H.A., Buck K.W.;
RT "Complete nucleotide sequence of the infectious cloned DNA components
RT of tomato golden mosaic virus: potential coding regions and regulatory
RT sequences."
RL EMBO J. 3:2197-2205 (1984).
CC -!- SIMILARITY: BELONGS TO GEMINIVIRUSES BL1 PROTEIN FAMILY.
CC -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 173
CC ONWARD DUE TO A FRAMESHIFT ERROR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M73794; AAA46584.1; -
CC EMBL; K02030; -; NOT_ANNOTATED_CDS.
CC PIR; A04169; QQCVLG.
CC InterPro; IPR000211; Gemin1.BL.
CC Pfam; PF00845; Gemin1.BL; 1.
CC CONFLICT 11 11 V -> A (IN REF. 2).
SQ SEQUENCE 293 AA; 32933 MW; E024E3F285C89012 CRC64;

```

```

Query Match 0.8%; Score 8; DB 1; Length 293;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 177 NYIESNRD 184
DB 13 NYIESNRD 20

RESULT 5
TR6B_HUMAN
ID TR6B_HUMAN STANDARD; PRT; 300 AA.
AC O95407;
DC 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tumor necrosis factor receptor superfamily member 6B precursor (Decoy
DE receptor for Fas ligand) (Decoy receptor 3) (DCR3) (M68).
GN TNFRSF6B OR DCR3 OR TR6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal lung; PubMed=9872321;
RX MEDLINE=99087326; PubMed=9872321;
RA Pitti R.M., Marsters S.A., Lawrence D.A., Roy M., Kischkel F.C.,
RA Dowd P., Huang A., Donahue C.J., Sherwood S.W., Baldwin D.T.,
RA Godowski P.J., Wood W.I., Gurney A.L., Hillan K.J., Cohen R.L.,
RA Goddard A.D., Botstein D., Ashkenazi A.;
RT "Genomic amplification of a decoy receptor for Fas ligand in lung and
RT colon cancer."
RL Nature 396:699-703 (1998).
RN [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 30-35.
RC TISSUE=Prostate;
RX MEDLINE=99253915; PubMed=10318773;
RA Yu K.-Y., Kwon B., Ni J., Zhai Y., Ebner R., Kwon B.S.;
RT "A newly identified member of tumor necrosis factor receptor
RT superfamily (TR6) suppresses LIGHT-mediated apoptosis."
RL J. Biol. Chem. 274:13733-13736 (1999).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=20122600; PubMed=10655513;
RA Bai C., Connolly B., Metzker M.L., Hilliard C.A., Liu X., Sandig V.,
RA Soderman A., Galloway S.M., Liu Q., Austin C.P., Caskey C.T.;
RT "Overexpression of M68/Dcr3 in human gastrointestinal tract tumors
RT independent of gene amplification and its location in a four-gene
RT cluster."
RL Proc. Natl. Acad. Sci. U.S.A. 97:1230-1235 (2000).
RN [4]
RP SEQUENCE FROM N.A.
RA Matthews L.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Strausberg R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Decoy receptor for the cytotoxic ligands TNFSF14/LIGHT
CC and TNFRSF6/FasL. Protects against apoptosis.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Detected in fetal lung, brain and liver.
CC Detected in adult stomach, spinal cord, lymph node, trachea,
CC spleen, colon and lung. Highly expressed in several primary tumors
CC from colon, stomach, rectum, esophagus and in SW480 colon
CC carcinoma cells.
CC -!- SIMILARITY: CONTAINS 4 TNFR-CYS REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC use the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF104419; AAD03056.1; -
DR EMBL; AF134240; AAD29688.1; -
DR EMBL; AF217796; AAF35244.1; -
DR EMBL; AF217793; AAF33685.1; -
DR EMBL; AF217794; AAF33686.1; -
DR EMBL; AL121845; CAC03668.1; -
DR EMBL; BC017065; AAH17065.1; -
DR Genew; HGNC:11921; TNFRSF6B.
DR MIM; 603361; -
DR HSSP; O14763; ID0G.
DR InterPro; IPR001368; TNFR_c6.
DR Pfam; PF00020; TNFR_c6; 4.
DR ProDom; PD000771; TNFR_c6; 1.
DR SMART; SM00208; TNFR_3.
DR PROSITE; PS00652; TNFR_NGFR_1; 2.
DR PROSITE; PS50050; TNFR_NGFR_2; 2.
KW Receptor; Apoptosis; Glycoprotein; Repeat; Signal.
FT SIGNAL 1 29
FT CHAIN 30 300 TUMOR NECROSIS FACTOR RECEPTOR
FT SUPERFAMILY MEMBER 6B.
FT REPEAT 31 70 TNFR-CYS 1.
FT REPEAT 72 113 TNFR-CYS 2.
FT REPEAT 115 150 TNFR-CYS 3.
FT REPEAT 152 193 TNFR-CYS 4.
FT DISULFID 49 62 BY SIMILARITY.
FT DISULFID 52 70 BY SIMILARITY.
FT DISULFID 73 88 BY SIMILARITY.
FT DISULFID 91 105 BY SIMILARITY.
FT DISULFID 95 113 BY SIMILARITY.
FT DISULFID 115 126 BY SIMILARITY.
FT DISULFID 132 150 BY SIMILARITY.
FT DISULFID 153 168 BY SIMILARITY.
FT DISULFID 174 193 BY SIMILARITY.
FT CARBOHYD 173 173 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 300 AA; 32679 MW; F90AEE33718449AF CRC64;

Query Match 0.8%; Score 8; DB 1; Length 300;
Best Local Similarity 100.0%; Pred. No. 5.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 410 PCPPGTFS 417
| | | | | | | |
DB 152 PCPPGTFS 159

RESULT 6
DCP2_PEA
ID DCP2_PEA STANDARD; PRT; 405 AA.
AC P51851;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pyruvate decarboxylase isozyme 2 (EC 4.1.1.1) (PDC) (Fragment).
GN PDC2.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
OX NCBI_TaxID=3888;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Miko;
RX MEDLINE=96215432; PubMed=8647075;
RA Muecke U., Wohlfarth T., Fiedler U., Baumlein H.,
RA Ruecknagel K.P., Koenig S.;
RA "Pyruvate decarboxylase from Pisum sativum. Properties, nucleotide

RT and amino acid sequences.";
RL Eur. J. Biochem. 237:373-382(1996).
CC -!- CATALYTIC ACTIVITY: A 2-oxo acid - an aldehyde + CO(2).
CC -!- COFACTOR: Binds 1 thiamine pyrophosphate and 1 metal ion per
CC subunit.
CC -!- SUBUNIT: HOMOOCTAMER (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE TPP ENZYMES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z66544; CAA91445.1; -
DR HSSP; P08672; IZPD.
DR InterPro; IPR000399; TPP_enzyme.
DR Pfam; PF00205; TPP_enzymes; 1.
DR Pfam; PF02775; TPP_enzymes_C; 1.
DR PROSITE; PS00187; TPP_ENZYMES; PARTIAL.
KW Lyase; Decarboxylase; Flavoprotein; Thiamine pyrophosphate;
KW Multigene family.
FT NON_TER 1
FT SIGNAL 1
SQ SEQUENCE 405 AA; 44078 MW; 44DDCE90B38677FB CRC64;

Query Match 0.8%; Score 8; DB 1; Length 405;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 KAVKPVLV 276
| | | | | | | |
DB 50 KAVKPVLV 57

RESULT 7
DCP1_TOBAC
ID DCP1_TOBAC STANDARD; PRT; 418 AA.
AC P51845;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pyruvate decarboxylase isozyme 1 (EC 4.1.1.1) (PDC) (Fragment).
GN PDC1.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Samsun; TISSUE=Leaf;
RX MEDLINE=95375236; PubMed=7647304;
RA Bucher M., Brander K., Sbicego S., Mandel T., Kuhlmeier C.;
RA "Aerobic fermentation in tobacco pollen.";
RL Plant Mol. Biol. 28:739-750(1995).
CC -!- CATALYTIC ACTIVITY: A 2-oxo acid - an aldehyde + CO(2).
CC -!- COFACTOR: Binds 1 thiamine pyrophosphate and 1 metal ion per
CC subunit.
CC -!- SUBUNIT: HOMOOCTAMER (POTENTIAL).
CC -!- TISSUE SPECIFICITY: LEAVES.
CC -!- INDUCTION: ANAEROBICALLY.
CC -!- SIMILARITY: BELONGS TO THE TPP ENZYMES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

```

DR EMBL; X81854; CAA57447.1; -.
DR HSP; P06169; IPVD.
DR InterPro; IPR00399; TPP_enzyme.
DR Pfam; PF00205; TPP_enzymes_1.
DR Pfam; PF02775; TPP_enzymes_C; 1.
DR Pfam; PF02776; TPP_enzymes_N; 1.
DR PROSITE; PS00187; TPP_ENZYMES; PARTIAL.
KW Lyase; Decarboxylase; Flavoprotein; Thiamine pyrophosphate;
KW Multigene family.
FT NON_TER 1
FT NON_TER 418
SQ SEQUENCE 418 AA; 45010 MW; 7C3C33DAA9FFD804 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 418;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 KAVKPVLV 276
DB 155 KAVKPVLV 162
IIIIIIII

RESULT 8
TRIT_HUMAN
ID TRIT_HUMAN STANDARD; PRT; 430 AA.
AC Q96924; Q96J01; Q9BUX7;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tumor necrosis factor receptor superfamily member TNFRSF19L precursor
DE (Receptor expressed in lymphoid tissues).
GN TNFRSF19L OR RELT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF N-TERMINUS.
RC TISSUE=Lymphoma;
RX MEDLINE=21213541; PubMed=11313261;
RA Sica G.L., Zhu G., Tamada K., Liu D., Ni J., Chen L.;
RT "RELT, a new member of the tumor necrosis factor receptor superfamily,
RT is selectively expressed in hematopoietic tissues and activates
RT transcription factor NF-kappaB.";
RL Blood 97:2702-2707(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Retinoblastoma;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Salto K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NED0 human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon, and Eye;
RA Strausberg R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 121-430 FROM N.A.
RC TISSUE=Spleen;
RA Jikuya H., Takano J., Nomura N., Kikuno R., Nagase T., Ohara O.;
RT "The nucleotide sequence of a long cDNA clone isolated from human
RT spleen.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Mediates activation of NF-kappa-B. May play a role in T-
CC cell activation.
CC -!- SUBUNIT: Associates with TRAF1.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Probable).

```

```

CC -!- TISSUE SPECIFICITY: Highest levels are in spleen, lymph node,
CC thymus, peripheral blood leukocytes, bone marrow and fetal liver.
CC Very low levels in skeletal muscle, testis and colon. Not detected
CC in brain, kidney and pancreas.
CC -!- SIMILARITY: CONTAINS 1 TNFR-CYS REPEAT.
CC -!- CAUTION: Ref.4 sequence differs from that shown due to several
CC frameshifts.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF319553; AAK77356.1; -.
CC EMBL; AK027899; BAB5441.1; -.
CC EMBL; BC001812; AAH01812.1; -.
CC EMBL; BC017279; AAH17279.1; -.
CC EMBL; AK074128; BAB84954.1; ALT_FRAME.
CC Genew; HGNC:13764; TNFRSF19L.
CC PROSITE; PS00652; TNFR_NGFR_1; FALSE_NEG.
CC PROSITE; PS00500; TNFR_NGFR_2; FALSE_NEG.
CC Receptor; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 26
FT CHAIN 27 430
FT TUMOR NECROSIS FACTOR RECEPTOR
FT SUPERFAMILY MEMBER TNFRSF19L.
FT EXTRACELLULAR (POTENTIAL).
FT POTENTIAL.
FT DOMAIN 27 162
FT TRANSMEM 163 183
FT DOMAIN 184 430
FT REPEAT 50 90
FT CYTOPLASMIC (POTENTIAL).
FT TNFR-CYS.
FT DISULFID 51 65
FT BY SIMILARITY.
FT DISULFID 71 90
FT CARBOHYD 149 149
FT CONFLICT 122 122
FT CONFLICT 187 187
FT CONFLICT 273 273
FT CONFLICT 379 380
FT DL -> TR (IN REF. 3; AAH01812).
SQ SEQUENCE 430 AA; 46092 MW; 4A5AB9AE32D36101 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 430;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 410 PCPPGTFS 417
DB 50 PCPPGTFS 57
IIIIIIII

RESULT 9
TRIT_MACFA
ID TRIT_MACFA STANDARD; PRT; 430 AA.
AC Q9N092;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tumor necrosis factor receptor superfamily member TNFRSF19L precursor
DE (Receptor expressed in lymphoid tissues).
GN TNFRSF19L OR RELT.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC Cercopitheidae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=21458551; PubMed=11574149;
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirata M., Suto Y.,
RA Hirai M., Terao K., Suzuki Y., Sugano S., Hashimoto K., Kusuda J.;
RT "Assignment of 118 novel cDNAs of cynomolgus monkey brain to human
RT chromosomes.";

```

RL Gene 275:31-37(2001).
CC -!- FUNCTION: Mediates activation of NF-kappa-B (By similarity). May
CC play a role in T-cell activation.
CC -!- SUBUNIT: Associates with TRAF1 (By similarity).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Probable).
CC -!- SIMILARITY: CONTAINS 1 TNFR-CYS REPEAT.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB045039; BAB01621.1; -
DR InterPro: IPR001368; TNFR_C6.
DR PROSITE; PS00652; TNFR_NGFR_1; FALSE_NEG.
DR PROSITE; PS005050; TNFR_NGFR_2; FALSE_NEG.
DR SMART; SM00208; TNFR; 1.
KW Receptor; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 26
FT CHAIN 27 430
FT TUMOR NECROSIS FACTOR RECEPTOR
FT SUPERFAMILY MEMBER TNFRSF19L.
FT EXTRACELLULAR (POTENTIAL).
FT POTENTIAL.
FT DOMAIN 27 162
FT TRANSMEM 163 183
FT DOMAIN 184 430
FT REPEAT 50 90
FT CYTOPLASMIC (POTENTIAL).
FT TNFR-CYS.
FT DISULFID 51 65
FT BY SIMILARITY.
FT DISULFID 71 90
FT BY SIMILARITY.
FT CARBOHYD 149 149
FT N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 430 AA; 45850 MW; BA8DE92593E1E859 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 430;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 410 PCPPGTFS 417
Db 50 PCPPGTFS 57
|||||||
RESULT 10
DCP3_ORYSA
ID DCP3_ORYSA STANDARD; PRT; 585 AA.
AC P51849;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pyruvate decarboxylase isozyme 3 (EC 4.1.1.1) (PDC).
GN PDC3.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Indica-IR54; TISSUE=Callus;
RX MEDLINE=95148752; PubMed=7846174;
RA Hossain M.A., McGee J.D., Grover A., Dennis E., Peacock W.J.,
RA Hodges T.K.;
RT "Nucleotide sequence of a rice genomic pyruvate decarboxylase gene
RT that lacks introns: a pseudo-gene";
RL Plant Physiol. 106:1697-1698(1994).
CC -!- CATALYTIC ACTIVITY: A 2-oxo acid - an aldehyde + CO(2).
CC -!- COFACTOR: Binds 1 thiamine pyrophosphate and 1 metal ion per
CC subunit.
CC -!- SUBUNIT: HOMOOCTAMER (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE TPP ENZYMES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U27350; AAA90948.1; -
DR EMBL; U38199; AAB40530.1; -
DR HSSP; P06169; LPVD.
DR InterPro: IPR000399; TPP_enzyme.
DR Pfam; PF02025; TPP_enzymes_C; 1.
DR Pfam; PF02775; TPP_enzymes_N; 1.
DR PROSITE; PS00187; TPP_ENZYMES; FALSE_NEG.
KW Lyase; Decarboxylase; Flavoprotein; Thiamine pyrophosphate;
KW Multigene family.
FT ACT_SITE 89 89 BY SIMILARITY.

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U07338; AAA68289.1; -
DR HSSP; P06169; LPVD.
DR InterPro: IPR000399; TPP_enzyme.
DR Pfam; PF02025; TPP_enzymes_C; 1.
DR Pfam; PF02775; TPP_enzymes_N; 1.
DR PROSITE; PS00187; TPP_ENZYMES; FALSE_NEG.
KW Lyase; Decarboxylase; Flavoprotein; Thiamine pyrophosphate;
KW Multigene family.
FT ACT_SITE 71 71 BY SIMILARITY.
SQ SEQUENCE 585 AA; 62456 MW; F26AAAFAS8AE52 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 585;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 269 KAVKPVLY 276
Db 231 KAVKPVLY 238
|||||||
RESULT 11
DCP2_ORYSA
ID DCP2_ORYSA STANDARD; PRT; 603 AA.
AC P51848;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Pyruvate decarboxylase isozyme 2 (EC 4.1.1.1) (PDC).
GN PDC2.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Indica-IR54; TISSUE=Callus;
RX Huq M.E., Hossain M.A., Hodges T.K.;
RT "Cloning and sequencing of a cDNA encoding pyruvate decarboxylase 2
RT gene from rice";
RL (In) Plant Gene Register PGR95-072.
CC -!- CATALYTIC ACTIVITY: A 2-oxo acid - an aldehyde + CO(2).
CC -!- COFACTOR: Binds 1 thiamine pyrophosphate and 1 metal ion per
CC subunit.
CC -!- SUBUNIT: HOMOOCTAMER (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE TPP ENZYMES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U27350; AAA90948.1; -
DR EMBL; U38199; AAB40530.1; -
DR HSSP; P06169; LPVD.
DR InterPro: IPR000399; TPP_enzyme.
DR Pfam; PF02025; TPP_enzymes_C; 1.
DR Pfam; PF02775; TPP_enzymes_N; 1.
DR PROSITE; PS00187; TPP_ENZYMES; FALSE_NEG.
KW Lyase; Decarboxylase; Flavoprotein; Thiamine pyrophosphate;
KW Multigene family.
FT ACT_SITE 89 89 BY SIMILARITY.

SQ SEQUENCE 603 AA; 64301 MW; 65B83A6468DCC2EB CRC64;

Query Match 0.8%; Score 8; DB 1; Length 603;
Best Local Similarity 100.0%; Pred. No. 9.4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 269 KAVKPVLV 276
|||||

Db 249 KAVKPVLV 256

RESULT 12

KRP2_RAT ID KRP2_RAT STANDARD; PRT; 671 AA.

AC Q62909;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Kinesin-related protein 2.
GN KRP2;

OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Sprague-Dawley; TISSUE=Testis;

RX MEDLINE=96228687; PubMed=8688559;

RA Sperry A.O.; Zhao L.-P.;

RT "Kinesin-related proteins in the mammalian testes: candidate motors
for meiosis and morphogenesis.";

RL Mol. Biol. Cell 7:289-305(1996).

RN [2]

RP REVISIONS.

RA Sperry A.O.;

RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: MOTOR PROTEIN, ACTIVE IN MEIOSIS.

CC -!- TISSUE SPECIFICITY: TESTIS. LOCALIZED TO THE MEIOGONIA ACTIVE
CELLS OF THE SEMINIFEROUS EPITHELIA IN THE TESTIS.

CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. MCAK/KIF2
SUBFAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; U44979; AAC53528.1; -

CC HSSP; P17119; 3KAR.

DR InterPro; IPR001752; kinesin_motor.

DR Pfam; PF00225; kinesin; 1

DR PRINTS; PF00380; KINESINHEAVY.

DR SMART; SM00129; KISC; 1.

DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.

DR PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.

KW Motor protein; Microtubules; ATP-binding; Coiled coil; Meiosis.

FT DOMAIN 1 198 GLOBULAR (POTENTIAL).

FT DOMAIN 199 540 KINESIN-MOTOR (BY SIMILARITY).

FT DOMAIN 541 671 COILED COIL (POTENTIAL).

FT NP_BIND 294 301 ATP (POTENTIAL).

SQ SEQUENCE 671 AA; 75661 MW; F2B54598C78DE8E CRC64;
Query Match 0.8%; Score 8; DB 1; Length 671;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 846 VISVPSKC 853

|||||

Db 226 VISVPSKC 233

RESULT 13

MCAK_CRIGR ID MCAK_CRIGR STANDARD; PRT; 718 AA.

AC P70096;

DT 01-NOV-1997 (Rel. 35, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Mitotic centromere-associated kinesin (MCAK) (Kinesin-like protein 6).
GN KNSL6.

OS Cricetus griseus (Chinese hamster).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;

OC Cricetus.

OX NCBI_TaxID=10029;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=95122643; PubMed=7822426;

RA Wordeman L.; Mitchison T.J.;

RT "Identification and partial characterization of mitotic centromere-
associated kinesin, a kinesin-related protein that associates with
centromeres during mitosis.";

RL J. Cell Biol. 128:95-105(1995).

RN [2]

RP REVISIONS.

RA Wordeman L.;

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: PRESENT THROUGHOUT THE CELL CYCLE, ASSOCIATES WITH
CENTROMERES AT EARLY PROPHASE, AND REMAINS ASSOCIATED WITH THE

CC CENTROMERE UNTIL AFTER TELOPHASE.

CC -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear.

CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. MCAK/KIF2
SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; U11790; AAB17358.2; -

CC HSSP; P17119; 3KAR.

DR InterPro; IPR001752; kinesin_motor.

DR Pfam; PF00225; kinesin; 1.

DR PRINTS; PF00380; KINESINHEAVY.

DR SMART; SM00129; KISC; 1.

DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.

DR PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.

KW Motor protein; Microtubules; ATP-binding; Coiled coil;

FT DOMAIN 1 246 GLOBULAR (POTENTIAL).

FT DOMAIN 247 612 KINESIN-MOTOR (BY SIMILARITY).

FT DOMAIN 613 651 COILED COIL (POTENTIAL).

FT DOMAIN 689 716 COILED COIL (POTENTIAL).

FT NP_BIND 342 349 ATP (POTENTIAL).

FT DOMAIN 409 412 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).

SQ SEQUENCE 718 AA; 80918 MW; 16ABDB8C66AD11B2 CRC64;

Query Match 0.8%; Score 8; DB 1; Length 718;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 846 VISVPSKC 853

|||||

Db 274 VISVPSKC 281

RESULT 14

TAT_SIVCZ

ID TAT_SIVCZ

AC P17285; STANDARD; PRT; 100 AA.

01-AUG-1990 (Rel. 15, Created)
01-AUG-1990 (Rel. 15, Last sequence update)
15-JUN-2002 (Rel. 41, Last annotation update)
TAT protein (Transactivating regulatory protein).
TAT.
GN
OS Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).
OC Viruses; Retroviridae; Retroviridae; Lentivirus.
OX NCBI_TaxID=11723;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90259077; PubMed=2188136;
RA Huet T., Cheynier R., Meyerhans A., Roelants G., Wain-Hobson S.;
RT "Genetic organization of a chimpanzee lentivirus related to HIV-1.";
RL Nature 345:356-359(1990).
CC -1- FUNCTION: TRANSCRIPTIONAL REGULATOR THAT ACTS BY BINDING TO THE
CC TRANS-ACTIVATING RESPONSIVE SEQUENCE (TAR) RNA ELEMENT AND
CC ACTIVATES TRANSCRIPTION INITIATION AND/OR ELONGATION FROM THE LTR
CC PROMOTER.
CC -1- SUBUNIT: BINDS CYCLIN T1 (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52154; CAA36404.1; -
DR PIR; S09987; TNJCSI.
DR HIV; X52154; TATSCPZ.
DR InterPro: IPR001831; HIV_Tat.
DR Pfam; PF00539; Tat; 1.
DR PRINTS; PR00055; HIVTATDOMAIN.
KW Transcription regulation; Activator; RNA-binding; Nuclear protein;
KW AIDS.
KW SEQUENCE 100 AA; 11209 MW; 1B78830B90EED50E CRC64;
SQ
Query Match 0.7%; Score 7; DB 1; Length 100;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 658 GSKNOD 664
Db [1]
Db 62 GSKNOD 68
RESULT 15
ID URE3_ECOLI STANDARD; PRT; 100 AA.
AC Q03282;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Urease gamma subunit (EC 3.5.1.5) (Urea amidohydrolase).
GN UREA.
OS Escherichia coli.
OG Plasmid.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93194816; PubMed=8449894;
RA D'Orazio S.E., Collins C.M.;
RT "Characterization of a plasmid-encoded urease gene cluster found in
RT members of the family Enterobacteriaceae.";
RL J. Bacteriol. 175:1860-1864(1993).
CC -1- CATALYTIC ACTIVITY: Urea + H(2)O -> CO(2) + 2 NH(3).
CC -1- SUBUNIT: (ALPHA, BETA, GAMMA)(3) (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE UREASE GAMMA SUBUNIT FAMILY.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L03307; AAA24745.1; -
DR PIR; B47090; B47090.
DR HSP; P18316; IFWB.
DR InterPro: IPR002026; Urease_gamma.
DR Pfam; PF00547; urease_gamma; 1.
DR ProDom; PD002319; Urease_gamma; 1.
DR TIGRFAMs; TIGR00193; urease_gamma; 1.
KW Hydrolase; Plasmid.
KW SEQUENCE 100 AA; 10983 MW; FBFD74F4C86B0805 CRC64;
SQ
Query Match 0.7%; Score 7; DB 1; Length 100;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 977 SCATMEG 983
Db [1]
Db 40 SCATMEG 46
Search completed: May 12, 2003, 13:39:37
Job time : 16.9342 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 13:30:09 ; Search time 33.5452 Seconds

(without alignments)
6308.212 Million cell updates/sec

Title: US-10-073-333A-4

Perfect score: 1027

Sequence: 1 MLFRAGFVRGNGRPAEA.....KEKEDHFESVOLKTSRSPNI 1027

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL_21.*

- 1: sp.archaea.*
- 2: sp.bacteria.*
- 3: sp.fungi.*
- 4: sp.human.*
- 5: sp.invertebrate.*
- 6: sp.mammal.*
- 7: sp.mhc.*
- 8: sp.organelle.*
- 9: sp.phage.*
- 10: sp.plant.*
- 11: sp.podent.*
- 12: sp.virus.*
- 13: sp.vertebrate.*
- 14: sp.unclassified.*
- 15: sp.rvirus.*
- 16: sp.bacteriap.*
- 17: sp.archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	342	33.3	493	4	Q96DP2
2	21	2.0	300	11	Q8R215
3	21	2.0	580	4	Q9P2W2
4	9	0.9	552	10	Q82444
5	8	0.8	55	12	Q67608
6	8	0.8	148	17	Q9HR77
7	8	0.8	154	12	Q9G325
8	8	0.8	154	12	Q9EGT5
9	8	0.8	239	16	Q9ABT4
10	8	0.8	268	16	Q926W3
11	8	0.8	268	16	Q8Y3P0
12	8	0.8	281	5	Q9TZN3
13	8	0.8	286	16	Q9PR14
14	8	0.8	293	12	Q8V9I1
15	8	0.8	379	2	Q50198
16	8	0.8	380	8	Q9TFR5

17	8	0.8	380	8	Q958X6
18	8	0.8	428	16	Q8ZF14
19	8	0.8	444	4	Q23079
20	8	0.8	447	5	O44593
21	8	0.8	498	4	O43267
22	8	0.8	550	4	Q9P201
23	8	0.8	593	10	Q8RUU6
24	8	0.8	597	10	Q9LGL5
25	8	0.8	605	10	Q8S4W8
26	8	0.8	606	10	Q8S4W9
27	8	0.8	607	10	Q9FFT4
28	8	0.8	607	10	Q9G536
29	8	0.8	622	2	Q45153
30	8	0.8	704	10	Q04434
31	8	0.8	721	11	Q92258
32	8	0.8	742	16	Q8YN26
33	8	0.8	881	10	Q65924
34	8	0.8	1051	12	Q91EU4
35	8	0.8	1059	10	Q9LH54
36	8	0.8	1206	16	Q9CCX9
37	8	0.8	1232	10	Q9LJ01
38	8	0.8	1578	16	Q92E25
39	8	0.8	1582	16	Q8Y9A5
40	8	0.8	2712	16	Q9F3X5
41	8	0.8	3102	5	Q9TZR4
42	8	0.8	3503	5	Q24292
43	7	0.7	65	6	Q95J37
44	7	0.7	79	2	Q9AFQ6
45	7	0.7	84	11	Q8VIN6

ALIGNMENTS

RESULT 1

Q96DP2 ID Q96DP2 PRELIMINARY; PRT; 493 AA.
AC Q96DP2;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE CDNA FLJ31340 fis, clone MESAN1000035, weakly similar to major surface-labeled trophozoite antigen precursor.
DE Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuko Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK055902; BAB71041.1;
DR InterPro; IPR001881; EGF_Ca.
DR SMART; SM00179; EGF_CA_1;
SQ SEQUENCE 493 AA: 54585 MW: 0AEBB2C19709B211 CRC64;

Query Match 33.3%; Score 342; DB 4; Length 493;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 342; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 334 GSSECTERPPCTTKDYFOIHPFCDDEGKTQIMYKWIPIKICREDLTDAIRLPPSGKKDC 393

Db 96 GSSECTERPPCTTKDYFOIHPFCDDEGKTQIMYKWIPIKICREDLTDAIRLPPSGKKDC 155

Qy 394 PPCNPGFYNNSSSCHPCPPGTFSDGTRECRPCPAGTEPALGFYKWNVLPGNKKTSFCF 453

```
Db 156 PPCNPGFYNNSSSCHPCPGTFSGTKECRPCAGTEPALGFEYKWNVLPGNNKTSCT 215
Qy 454 NVGNSKCDGMNGWEVAGDHIOGAGGSDNDYLILNLHPIPGFKPPTSMGTGATGSELGRITF 513
Db 216 NVGNSKCDGMNGWEVAGDHIOGAGGSDNDYLILNLHPIPGFKPPTSMGTGATGSELGRITF 275
Qy 514 VFETLCSADCVLYFMVDINRKSTNNVSWGGTKKQAVTHIIFKNATFTTWFQFRTNOG 573
Db 276 VFETLCSADCVLYFMVDINRKSTNNVSWGGTKKQAVTHIIFKNATFTTWFQFRTNOG 335
Qy 574 QNRRFINDMKIYSITATNAVGVASSCRACALGSESGSCVPCPGPHYIEKETNOCK 633
Db 336 QNRRFINDMKIYSITATNAVGVASSCRACALGSESGSCVPCPGPHYIEKETNOCK 395
Qy 634 ECPPTDYLISHOVYKEACIPCGPSKNNQDHSVCYSOCFFY 675
Db 396 ECPPTDYLISHOVYKEACIPCGPSKNNQDHSVCYSOCFFY 437
RESULT 2
Q8R215 PRELIMINARY; PRT; 300 AA.
AC Q8R215:
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Hypothetical 32.6 kDa protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Strausberg R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC022655; AAB22655.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 300 AA; 32642 MW; 6455109C054B6CB4 CRC64;
Query Match 2.0%; Score 21; DB 11; Length 300;
Best Local Similarity 100.0%; Pred. No. 4e-13;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 943 LTCYFVKKNQKLEYKSKLVM 963
Db 213 LTCYFVKKNQKLEYKSKLVM 233
RESULT 3
Q9P2M2 PRELIMINARY; PRT; 580 AA.
AC Q9P2M2:
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE KIAA1324 protein (Fragment).
GN KIAA1324.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=20181126; PubMed=10718198;
RA Nagase T., Kikuno R., Ishikawa K., Hirokawa M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XVI.
RT The complete sequences of 150 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:65-73(2000).
```

```
DR EMBL; AB037745; BAA92562.1; -.
FT NON_TER 1
SQ SEQUENCE 580 AA; 63365 MW; 5B63C19265EC8E3C CRC64;
Query Match 2.0%; Score 21; DB 4; Length 580;
Best Local Similarity 100.0%; Pred. No. 7.1e-13;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 943 LTCYFVKKNQKLEYKSKLVM 963
Db 502 LTCYFVKKNQKLEYKSKLVM 522
RESULT 4
Q82444 PRELIMINARY; PRT; 552 AA.
AC Q82444:
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Peroxisomal targeting sequence 1 receptor (Fragment).
GN PEX5.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99007315; PubMed=9789089;
RA Kragler F., Lametschwandner G., Christmann J., Hartig A.,
RA Harada J.J.;
RT "Identification and analysis of the plant peroxisomal targeting signal
RT 1 receptor NPEX5.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:13336-13341(1998).
DR EMBL; AF056282; AAC69180.1; -.
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 4.
DR SMART; SM00028; TPR; 4.
KW Receptor.
FT NON_TER 1
SQ SEQUENCE 552 AA; 62293 MW; 7045FA177B0F51C6 CRC64;
Query Match 0.9%; Score 9; DB 10; Length 552;
Best Local Similarity 100.0%; Pred. No. 3.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 4 RARGPVRGR 12
Db 22 RARGPVRGR 30
RESULT 5
Q67608 PRELIMINARY; PRT; 55 AA.
AC Q67608:
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Tomato golden mosaic virus subgenomic DNA derived from DNA B ccdds -
DE covalently closed circular double-stranded molecule.
OS Tomato golden mosaic virus (TGMV).
OC Viruses; ssDNA viruses; Geminiviridae; Begomovirus.
OX NCBI_TaxID=10831;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87040767; PubMed=3022243;
RA MacDowell S.W., Coutts R.H.A., Buck K.W.;
RT "Molecular characterization of subgenomic single-stranded and double-
RT stranded DNA forms isolated from plants infected with tomato golden
RT mosaic virus.";
RL Nucleic Acids Res. 14:7967-7984(1986).
DR EMBL; X04485; CAA28171.1; -.
```

DR InterPro: IPR000211; Gemini.BL.
 DR Pfam: PF00845; Gemini.BL1; 1.
 SQ SEQUENCE 55 AA; 6281 MW; 41963570D739EC96 CRC64;

Query Match 0.8%; Score 8; DB 12; Length 55;
 Best Local Similarity 100.0%; Pred. No. 5.4;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 177 NVIESNRD 184
 |||||

DB 13 NVIESNRD 20

RESULT 6

Q9HR77 PRELIMINARY; PRT; 148 AA.

AC Q9HR77; (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Vng0825C.
 GN VNG0825C.

OS Halobacterium sp. (strain NRC-1).
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
 OC Halobacteriaceae; Halobacterium.

OX NCBI_TaxID=64091;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20504483; PubMed=11016950;

RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
 RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
 RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
 RA Isenbarger T.A., Peck R.F., Pohlshroder M., Spudis J.L., Jung K.-H.,
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
 RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.,
 RT "Genome sequence of Halobacterium species NRC-1.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
 DR EMBL: AF005024; AAG19281.1; -;
 DR InterPro: IPR004435; MOBB.

DR Pfam: PF03205; MOBB; 1.

KW Complete proteome.

SQ SEQUENCE 148 AA; 15073 MW; F242336129C362FE CRC64;

Query Match 0.8%; Score 8; DB 17; Length 148;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 155 TVVGPSDS 162
 |||||

DB 5 TVVGPSDS 12

RESULT 7

Q9DGZ5 PRELIMINARY; PRT; 154 AA.

AC Q9DGZ5; (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE X protein.

GN X.

OS Hepatitis B virus.

OC Viruses; Retroviridae; Hepadnaviridae; Orthohepadnavirus.

OX NCBI_TaxID=10407;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=14118, AND 11141;

RX PubMed=10950984;

RA Hannoun C., Norder H., Lindh M.;

RT "An aberrant genotype revealed in recombinant hepatitis B virus

strains from Vietnam.";

RL J. Gen. Virol. 81:2267-2272(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=14118, AND 11141;
 RA Hannoun C., Norder H., Lindh M.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF241408; AAG17579.1; -;

DR EMBL: AF241407; AAG17572.1; -;

DR InterPro: IPR000236; TransactX.

DR Pfam: PF00739; X; 1.

SQ SEQUENCE 154 AA; 16516 MW; A13D1D623218640D CRC64;

Query Match

0.8%; Score 8; DB 12; Length 154;

Best Local Similarity 100.0%; Pred. No. 13;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 GDLPSSSS 55
 |||||

DB 35 GDLPSSSS 42

RESULT 8

Q9E6T5

ID Q9E6T5 PRELIMINARY; PRT; 154 AA.

AC Q9E6T5;

DT 01-MAR-2001 (TREMBlrel. 16, Created)

DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE X protein;

GN X.

OS Hepatitis B virus.

OC Viruses; Retroviridae; Hepadnaviridae; Orthohepadnavirus.

OX NCBI_TaxID=10407;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=6871;

RX MEDLINE=20409088; PubMed=10950984;

RA Hannoun C., Norder H., Lindh M.;

RT "An aberrant genotype revealed in recombinant hepatitis B virus

strains from Vietnam.";

RL J. Gen. Virol. 81:2267-2272(2000).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=6871;

RA Hannoun C., Norder H., Lindh M.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF241409; AAG17586.1; -;

DR InterPro: IPR000236; TransactX.

DR Pfam: PF00739; X; 1.

SQ SEQUENCE 154 AA; 16563 MW; E6B37B623218641B CRC64;

Query Match

0.8%; Score 8; DB 12; Length 154;

Best Local Similarity 100.0%; Pred. No. 13;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 GDLPSSSS 55
 |||||

DB 35 GDLPSSSS 42

RESULT 9

Q9ABT4

ID Q9ABT4 PRELIMINARY; PRT; 239 AA.

AC Q9ABT4;

DT 01-JUN-2001 (TREMBlrel. 17, Created)

DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)

DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)

DE Hypothetical protein CC0136.

GN CC0136.

OS Caulobacter crescentus.

OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;

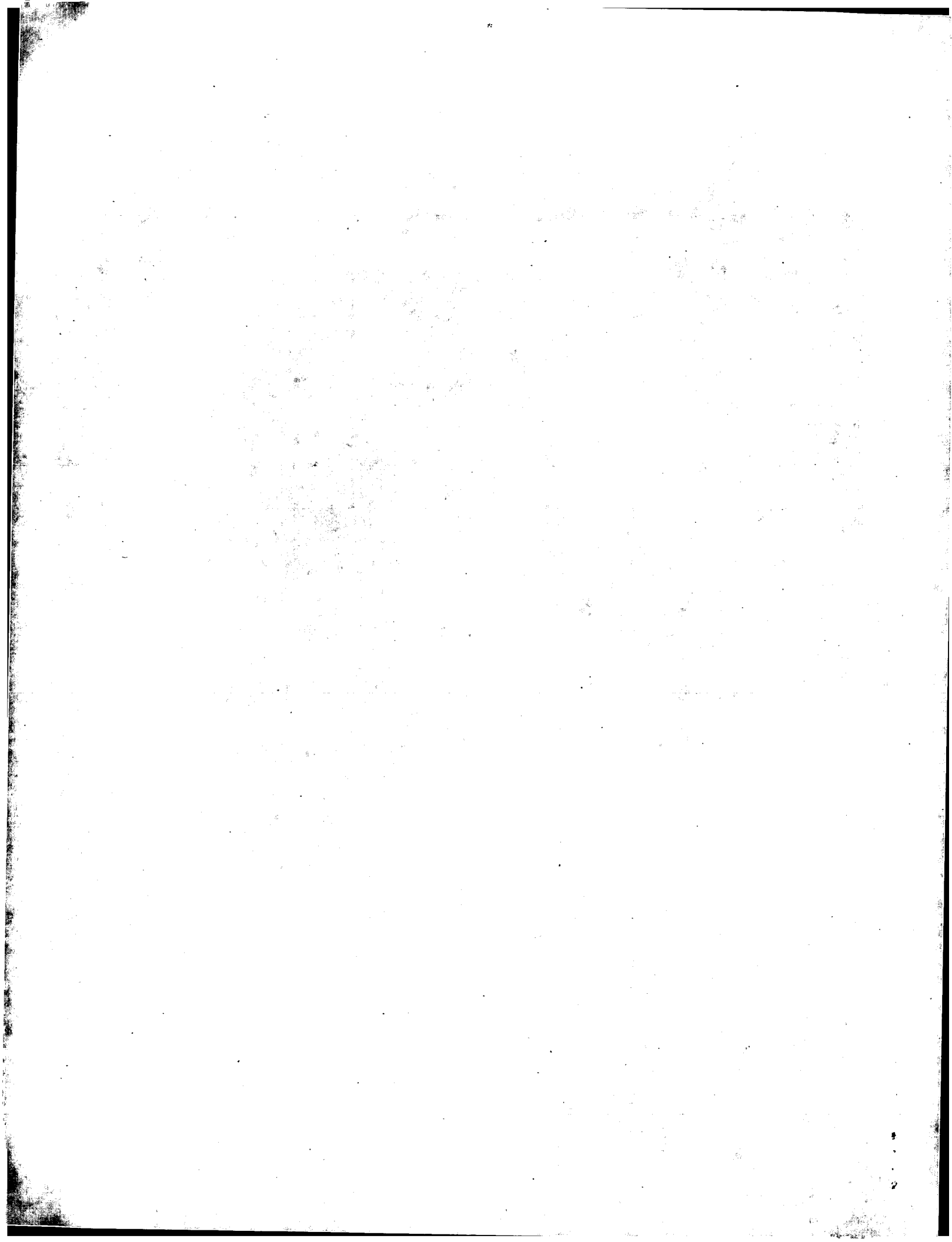
OC Caulobacter.

OX NCBI_TaxID=155892;

RESULT 11	Q8Y3P0	PRELIMINARY:	PRT:	268 AA.
ID	Q8Y3P0	PRELIMINARY:	PRT:	268 AA.
AC	Q8Y3P0:			
DT	01-WAR-2002 (TReMBLrel. 20, Created)			
DT	01-WAR-2002 (TReMBLrel. 20, Last sequence update)			
DT	01-JUN-2002 (TReMBLrel. 21, Last annotation update)			
DE	Hypothetical protein lmo2795.			
GN	lmo2795.			
OS	Listeria monocytogenes.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;			
OC	Listeriaceae; Listeria.			
OX	NCBI_TaxID=1639;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=EGD-E / SEROVAR 1/2A;			
RC	MEDLINE=2153779; PubMed=11679669;			
RA	Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,			
RA	Baquerio F., Berche P., Bloecker H., Brandt P., Chakraborty T.,			
RA	Charbit A., Chetoui F., Couve E., de Daruvar A., Dehoux P.,			
RA	Dommann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,			
RA	Entian K.-D., Fshih H., Garcia-del Portillo F., Garrido P.,			
RA	Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,			
RA	Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,			
RA	Madieno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,			
RA	Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,			
RA	Rommel B., Rose M., Schlueter T., Simoes N., Tierrez A.,			
RA	Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;			
RT	"Comparative genomics of Listeria species.";			
RL	Science 294:849-852(2001).			
RL	EMBL: AL591984; CAD01008.1.			
DR	ListIList; LMO02795; -			
DR	InterPro: IPR000281; HTH_RpIR.			
DR	InterPro: IPR001347; SIS.			
DR	Pfam: PF01418; HTH_6; 1.			
DR	Pfam: PF01380; SIS; 1.			
KW	Hypothetical protein; Complete genome.			
SEQUENCE	268 AA: 30643 MW; 7C26214E8FD1DAA2 CRC64;			
Query Match *	0.8%; Score 8; DB 16; Length 268;			
Best Local Similarity	100.0%; Pred. No. 22;			
Matches	8; Conservative 0; Mismatches 0; Indels 0; Gaps			
QY	792 LKNIKE 799			
DB	32 LKNIKE 39			
PRELIMINARY:	PRT:	281 AA.		
Q97ZN3	PRELIMINARY:	PRT:	281 AA.	
ID	Q97ZN3			
AC	Q97ZN3:			
DT	01-MAY-2000 (TReMBLrel. 13, Created)			
DT	01-MAY-2000 (TReMBLrel. 13, Last sequence update)			
DT	01-JUN-2001 (TReMBLrel. 17, Last annotation update)			
DE	C43H8.3 protein.			
GN	C43H8.3.			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;			
OC	Rhabditidae; Peloderinae; Caenorhabditis.			
OX	NCBI_TaxID=6239;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=BRISTOL N2;			
RC	MEDLINE=94150718; PubMed=7906398;			
RA	Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,			
RA	Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,			
RA	Craxton M., Dear S., Du Z., Durbin R., Favetto A., Fulcon L.,			
RA	Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnson L.,			
RA	Jones M., Kershaw J., McKinst J., Laister N., Latreille P.,			
RA	Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,			

RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Snelson N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierly-Wieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.:
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Tin-Wollam A.M.;
RT "The sequence of C. elegans cosmid C43H8.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF098499; AAC67398.1; -;
DR InterPro; IPR001791; Laminin_G..
DR Pfam; PF00054; Laminin_G; 1.
DR SMART; SM00282; LamG; 1.
SQ SEQUENCE 281 AA; 31708 MW; 5253CC040277343C CRC64;
Query Match 0.8%; Score 8; DB 5; Length 281;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 731 NITDFTVK 738
DB 111 NITDFTVK 118
RESULT 13
Q9PR14 PRELIMINARY; PRT; 286 AA.
ID Q9PR14
AC Q9PR14
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE ATP synthase gamma chain.
GN ATPG OR U0130.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Ureaplasma.
OX NCBI_TaxID=134821;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SEROVAR 3;
RX MEDLINE=20500219; PubMed=11048724;
RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
RA Cassell G.H.;
RT "The complete sequence of the mucosal pathogen Ureaplasma
RT urealyticum.";
RL Nature 407:757-762(2000).
DR EMBL; AE021114; AAF30536.1; -;
DR InterPro; IPR000131; ATPase_gamma.
DR Pfam; PF00231; ATP-synt; 1.
DR PRINTS; PR00126; ATPASEGAMMA.
DR TIGRFAMS; TIGR01146; ATPsyn_Flgamma; 1.
KW Complete proteome.
SQ SEQUENCE 286 AA; 32295 MW; C9AE278976AAE4F5 CRC64;
Query Match 0.8%; Score 8; DB 16; Length 286;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 739 EIVAGSD 746
DB 279 EIVAGSD 286
RESULT 14

Q8V9I1 PRELIMINARY; PRT; 293 AA.
ID Q8V9I1
AC Q8V9I1;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Movement protein.
GN MP.
OS Squash yellow mottle virus.
OC Viruses; ssDNA viruses; Geminiviridae; Begomovirus.
OX NCBI_TaxID=91372;
RN [1]
RP SEQUENCE FROM N.A.
RA Ramirez P., Karkashian J., Zuniga C., Maxwell D.;
RT "Association of Squash Yellow Mottle Virus with cucurbits and papaya
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF440790; AAL33677.1; -;
DR InterPro; IPR000211; Gemini_BL.
DR Pfam; PF00845; Gemini_BL; 1.
SQ SEQUENCE 293 AA; 33078 MW; 8F99918F4FFE54AC CRC64;
Query Match 0.8%; Score 8; DB 12; Length 293;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 177 NYIESNRD 184
DB 13 NYIESNRD 20
RESULT 15
Q50198 PRELIMINARY; PRT; 379 AA.
ID Q50198
AC Q50198
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE L222-ORF9.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97124199; PubMed=8969512;
RA Fsihi H., De Rossi E., Salazar L., Cantoni R., Labo M., Riccardi G.,
RA Takiff H.E., Eiglmeyer K., Bergh S., Cole S.T.;
RT "Gene arrangement and organization in a approximately 76 kb fragment
RT encompassing the oric region of the chromosome of Mycobacterium
RT leprae.";
RL Microbiology 142:0-0(0).
DR EMBL; L39923; AAB53129.1; -;
DR InterPro; IPR004268; MVIN_like.
DR Pfam; PF03023; MVIN; 1.
SQ SEQUENCE 379 AA; 40010 MW; FD23F9F20B4A74F5 CRC64;
Query Match 0.8%; Score 8; DB 2; Length 379;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 936 TAVLLVAL 943
DB 243 TAVLLVAL 250
Search completed: May 12, 2003, 13:40:54
Job time : 35.5452 secs



GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.
OM protein - protein search, using sw model
Run on: May 12, 2003, 07:48:38 ; Search time 49.5437 seconds
(without alignments)
2762.174 Million cell updates/sec

Title: US-10-073-333A-4
Perfect score: 5681
Sequence: 1 MLFRARGVPRGWRGPAEA.....KEKEDHFEVQLKTSRSPNI 1027

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.2
1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	5681	100.0	1027	22	AAB70256
2	5307	93.4	963	22	AAB70255
3	3050.5	53.7	1013	21	AAB70254
4	3040.5	53.5	1013	22	AAB83845
5	3025	53.2	1013	22	AAU12190
6	2982.5	52.5	1001	22	AAB35333
7	2712.5	47.7	911	22	AAB83850
8	2702.5	47.6	870	22	AAB83851
9	2526	44.5	464	22	AAB48377
10	2222	39.1	411	22	AAB48372

11	2167	38.1	750	22	AAB35328
12	1487.5	26.2	495	20	AAV59972
13	1208.5	21.3	383	22	AAB83853
14	883	15.5	372	22	AAB85768
15	705.5	12.4	209	22	AAB83852
16	452	8.0	81	22	ABB39918
17	452	8.0	81	22	ABB24471
18	452	8.0	81	22	ABM60663
19	452	8.0	81	22	AAW73335
20	452	8.0	81	22	AAW35335
21	444.5	7.8	208	21	ABG43186
22	444.5	7.8	208	21	ABG43186
23	444.5	7.8	208	21	ABG43186
24	395	7.0	78	22	AAW70281
25	370.5	6.5	147	22	AAW70285
26	353	6.2	60	22	ABB38686
27	353	6.2	60	22	ABB38686
28	353	6.2	60	22	ABB23759
29	353	6.2	60	22	AAW59318
30	353	6.2	60	22	AAW71867
31	353	6.2	60	22	AAW32149
32	353	6.2	60	23	ABG41680
33	353	6.2	64	22	AAW72925
34	353	6.2	105	21	ABG42760
35	350	6.2	150	20	AAV12274
36	322	5.7	71	22	AAU21345
37	255	4.5	50	22	ABB39681
38	255	4.5	50	22	ABB24346
39	255	4.5	50	22	AAW60397
40	255	4.5	50	22	AAW73033
41	255	4.5	50	22	AAW19811
42	255	4.5	50	22	AAW33257
43	255	4.5	50	23	ABG42877
44	205	3.6	1605	21	AAW19805
45	205	3.6	1605	21	AAW48454

ALIGNMENTS

RESULT 1
AAB70256
ID AAB70256 standard; protein; 1027 AA.
XX
AC AAB70256;
XX
DT 10-MAY-2001 (first entry)
XX
DE TR16-long receptor protein.
XX
TR16 recptor; tumour necrosis factor receptor superfamily;
KW apoptosis; inflammatory; cancer; immune; neurodegenerative.
XX
OS Unidentified.
XX
PN WO200112671-AL.
XX
PD 22-FEB-2001.
XX
PF 10-AUG-2000; 2000WO-US21885.
XX
PR 12-AUG-1999; 99US-0148348.
PR 13-AUG-1999; 99US-0148683.
PR 13-AUG-1999; 99US-0148870.
PR 16-AUG-1999; 99US-0148758.
PR 17-AUG-1999; 99US-0149181.
PR 18-AUG-1999; 99US-0149453.
PR 19-AUG-1999; 99US-0149498.
XX
(HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Young PE, Baker KP;
XX

DR WPI; 2001-138754/14.
XX New nucleic acid molecule encoding a TR16 tumor necrosis factor
PT receptor polypeptide, useful for the diagnosis and treatment of cancer,
PT autoimmune disorders and cardiovascular diseases -
XX
PS Disclosure; Fig 4; 286pp; English.
XX
CC The present invention relates to a TR16 receptor (tumour necrosis
CC factor receptor superfamily). The invention is useful treating
CC diseases and disorders associated with the inhibited or increased
CC apoptosis. In particular inflammatory diseases, cancers, immune and
CC neurodegenerative disorders may be treated.
XX
SQ Sequence 1027 AA;
Query Match 100.0%; Score 5681; DB 22; Length 1027;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1027; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLFRARGVRGWRGPAEAPRRGSPWPSPAWICWALAGCAAWAGDLPSSSRPLPP 60
DB 1 MLFRARGVRGWRGPAEAPRRGSPWPSPAWICWALAGCAAWAGDLPSSSRPLPP 60
QY 61 COEKDYHFEYTECDSSGSRWVAIPNSAVDCSGLPDPVRGKECTFCASGEYLEMKNQVC 120
DB 61 COEKDYHFEYTECDSSGSRWVAIPNSAVDCSGLPDPVRGKECTFCASGEYLEMKNQVC 120
QY 121 SKCGEGTSLGSGIKFDEWDELPAFESNATFMDTVVGPDSRDPGCCNNSWIPRGNYIE 180
DB 121 SKCGEGTSLGSGIKFDEWDELPAFESNATFMDTVVGPDSRDPGCCNNSWIPRGNYIE 180
QY 181 SNRDDCTSLIYAVHLKSGYFFQYVDNNIFEFQNDQOCQEMDTTDDKWKLTDN 240
DB 181 SNRDDCTSLIYAVHLKSGYFFQYVDNNIFEFQNDQOCQEMDTTDDKWKLTDN 240
QY 241 GEMGSHVLMKSGTNTLYWRTTILMGSKAVPVLVKNITIEGVAYTSECFCKPGTFSN 300
DB 241 GEMGSHVLMKSGTNTLYWRTTILMGSKAVPVLVKNITIEGVAYTSECFCKPGTFSN 300
QY 301 KPGSFNCQVCPNTYSEKAKCIRCKDQSGSSECTERPCTTKDYFOIHTPCDEEG 360
DB 301 KPGSFNCQVCPNTYSEKAKCIRCKDQSGSSECTERPCTTKDYFOIHTPCDEEG 360
QY 361 KTOIMYKWIPEKICREDLDAIRLPPSGEKKDCPCPCNPGFYNNSSCHPCPPGTFSDGT 420
DB 361 KTOIMYKWIPEKICREDLDAIRLPPSGEKKDCPCPCNPGFYNNSSCHPCPPGTFSDGT 420
QY 421 KECRCPAGTEPALGPEYKWNVLPONMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGS 480
DB 421 KECRCPAGTEPALGPEYKWNVLPONMKTSCFNVGNSKCDGMNGWEVAGDHIQSGAGGS 480
QY 481 DNDYLTLNLHPIGFKPPTSMGTATGSELGRITFVFETLCSADCLVLFMVDINRKSTNVVE 540
DB 481 DNDYLTLNLHPIGFKPPTSMGTATGSELGRITFVFETLCSADCLVLFMVDINRKSTNVVE 540
QY 541 SWGGTKEQAYTHIIFKNATFTTFAFORTNOQDNRREINDMKIYSITATNAVGVAS 600
DB 541 SWGGTKEQAYTHIIFKNATFTTFAFORTNOQDNRREINDMKIYSITATNAVGVAS 600
QY 601 SCRACALGSGQSSGVCPGPGHYIEKTNQCKECPDPTYLSIHQVYGKACI PCPGGSK 660
DB 601 SCRACALGSGQSSGVCPGPGHYIEKTNQCKECPDPTYLSIHQVYGKACI PCPGGSK 660
QY 661 NNDHVSVCYSDCFYHEKENQILHYDFSNLSSVGLMNGPSFTSKGTKYHFHFNISLCGH 720
DB 661 NNDHVSVCYSDCFYHEKENQILHYDFSNLSSVGLMNGPSFTSKGTKYHFHFNISLCGH 720
QY 721 EGKKMALCTNNITDFTVKEIVAGSDDDYTNLGVAFVQCSYTIIPSESKGFRALSSQSIIILA 780
DB 721 EGKKMALCTNNITDFTVKEIVAGSDDDYTNLGVAFVQCSYTIIPSESKGFRALSSQSIIILA 780
QY 781 DTFIGVTVETTLKNINIKEDMPVPPTSQIPDVHFFYKSSATTSCINGRSTAVKMRNCNPT 840

DB 781 DTFIGVTVETTLKNINIKEDMPVPPTSQIPDVHFFYKSSATTSCINGRSTAVKMRNCNPT 840
QY 841 KSGAGVISVPSKCPAGTCDGCTFYFLWESAEACPLCTEHDHFHEIGACKRGFQETLYVMN 900
DB 841 KSGAGVISVPSKCPAGTCDGCTFYFLWESAEACPLCTEHDHFHEIGACKRGFQETLYVMN 900
QY 901 EPKWCIKIGISLPEKKLATCETVDFWLKVGAGVGAFTAVLLVALTCYFWKKNOKLEYKYSK 960
DB 901 EPKWCIKIGISLPEKKLATCETVDFWLKVGAGVGAFTAVLLVALTCYFWKKNOKLEYKYSK 960
QY 961 LVMTTNSKECELPAAADSCAIMEGEDNEEBVYNSKQSLGKLSLATKEKEDHFEVSQVLK 1020
DB 961 LVMTTNSKECELPAAADSCAIMEGEDNEEBVYNSKQSLGKLSLATKEKEDHFEVSQVLK 1020
QY 1021 TSSRSPNI 1027
DB 1021 TSSRSPNI 1027
RESULT 2
AAB70255
ID AAB70255 standard; protein; 963 AA.
XX
XX AAB70255;
XX
XX 10-MAY-2001 (first entry)
XX
XX TR16-short receptor protein.
XX
XX TR16 receptor; tumour necrosis factor receptor superfamily;
KW apoptosis; inflammatory; cancer; immune; neurodegenerative.
XX
XX Unidentified.
XX
XX WO200112671-A1.
XX
XX 22-FEB-2001.
XX
XX 10-AUG-2000; 2000WO-US21885.
XX
XX 12-AUG-1999; 99US-0148348.
XX
XX 13-AUG-1999; 99US-0148683.
XX
XX 13-AUG-1999; 99US-0148870.
XX
XX 16-AUG-1999; 99US-0148758.
XX
XX 17-AUG-1999; 99US-0149181.
XX
XX 18-AUG-1999; 99US-0149453.
XX
XX 19-AUG-1999; 99US-0149498.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Young PE, Baker KP;
XX
XX WPI; 2001-138754/14.
XX
XX New nucleic acid molecule encoding a TR16 tumor necrosis factor
PT receptor polypeptide, useful for the diagnosis and treatment of cancer,
PT autoimmune disorders and cardiovascular diseases -
XX
XX Claim 1; Fig 1; 286pp; English.
XX
XX The present invention relates to a TR16 receptor (tumour necrosis
CC factor receptor superfamily). The invention is useful treating
CC diseases and disorders associated with the inhibited or increased
CC apoptosis. In particular inflammatory diseases, cancers, immune and
CC neurodegenerative disorders may be treated.
XX
SQ Sequence 963 AA;
Query Match 93.4%; Score 5307; DB 22; Length 963;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 953; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MLFRARGVGRGWGRPAEAPRRGRSPWPSPAWICWALAGCAQAAWAGDLPSSSSRPLPP 60
Db 1 MLFRARGVGRGWGRPAEAPRRGRSPWPSPAWICWALAGCAQAAWAGDLPSSSSRPLPP 60
QY 61 CQEKDYHFEYTECDSGSRWRVAIPNSAVDCSLDPLDVRGKECTFSCASGEYLEMKNOVC 120
Db 61 CQEKDYHFEYTECDSGSRWRVAIPNSAVDCSLDPLDVRGKECTFSCASGEYLEMKNOVC 120
QY 121 SKCGEGTSLGSGIKFDEWDELPAFNSIATFMDTVVGPDSRPGDCNNSWIPRCNYIE 180
Db 121 SKCGEGTSLGSGIKFDEWDELPAFNSIATFMDTVVGPDSRPGDCNNSWIPRCNYIE 180
QY 181 SNRDDCTVSLIYAVHLKKGSYFFVQYVDNIIFFEFFIONDQOQEMDITTDKWKVLTDN 240
Db 181 SNRDDCTVSLIYAVHLKKGSYFFVQYVDNIIFFEFFIONDQOQEMDITTDKWKVLTDN 240
QY 241 GEMGSHSVMLKSGTNTLYWRTTGILMGSKAVKPVLVKNITIEGVAYTSCFPCKPGTFSN 300
Db 241 GEMGSHSVMLKSGTNTLYWRTTGILMGSKAVKPVLVKNITIEGVAYTSCFPCKPGTFSN 300
QY 301 KPGSFNCQVCPNTYSEKAKECIRCKDDSQFSGSSECTERPPCTTKDYFOIHTPCDEEG 360
Db 301 KPGSFNCQVCPNTYSEKAKECIRCKDDSQFSGSSECTERPPCTTKDYFOIHTPCDEEG 360
QY 361 KTOIMYKWIPEKICREDLTDALRLPSGEKKDCPCPCNPGFYNNSSSCHPCPPGTFSGT 420
Db 361 KTOIMYKWIPEKICREDLTDALRLPSGEKKDCPCPCNPGFYNNSSSCHPCPPGTFSGT 420
QY 421 KECRCPAGTERPALGPEYKWNVLPNKMKTSCFNVGNSKCDGMNGWEVAGDHIQSCAGGS 480
Db 421 KECRCPAGTERPALGPEYKWNVLPNKMKTSCFNVGNSKCDGMNGWEVAGDHIQSCAGGS 480
QY 481 DNDYILNLHIHIFGFKPPTSMTGATSELGRITFVFETLCSADCVLYFMVDINRKSTNVYE 540
Db 481 DNDYILNLHIHIFGFKPPTSMTGATSELGRITFVFETLCSADCVLYFMVDINRKSTNVYE 540
QY 541 SWGGTKEKQAYTHIIFKNATFTFWAFORTNOGDNRRINDMWKIYSITATNAVGVAS 600
Db 541 SWGGTKEKQAYTHIIFKNATFTFWAFORTNOGDNRRINDMWKIYSITATNAVGVAS 600
QY 601 SCRACALGSEQSGSCVCPGPGHYIEKETNOCKECPPTDYLISHOVYKGEACIPCGPSK 660
Db 601 SCRACALGSEQSGSCVCPGPGHYIEKETNOCKECPPTDYLISHOVYKGEACIPCGPSK 660
QY 661 NNODHSVCYSDCFFHYEKENQILHDFSNLSSVGLMNGPSTSKGTYFHFENISLCGH 720
Db 661 NNODHSVCYSDCFFHYEKENQILHDFSNLSSVGLMNGPSTSKGTYFHFENISLCGH 720
QY 721 EGKKNALCTNNITDFTVKEIVAGSDDYTNLVGAFCVQSTIIPSEKGFRAALSSQSIILA 780
Db 721 EGKKNALCTNNITDFTVKEIVAGSDDYTNLVGAFCVQSTIIPSEKGFRAALSSQSIILA 780
QY 781 DFIGVTVETTLKNIKEDMPVPTSQIPDVHFFYKSTATTSCINGRSTAVKMRNPT 840
Db 781 DFIGVTVETTLKNIKEDMPVPTSQIPDVHFFYKSTATTSCINGRSTAVKMRNPT 840
QY 841 KSGAGVISVPKCPAGTCDGCTFYFLWESAECPLCTEHDHFHEIGACKRGFOETLYVWN 900
Db 841 KSGAGVISVPKCPAGTCDGCTFYFLWESAECPLCTEHDHFHEIGACKRGFOETLYVWN 900
QY 901 EPKWCIGISLPEKKLATCETVDFWLKYGAGVGAFTAVLLVALTCYFWKKNOK 953
Db 901 EPKWCIGISLPEKKLATCETVDFWLKYGAGVGAFTAVLLVALTCYFWKKNOK 953
RESULT 3
AAB26179
ID AAB26179 standard; Protein; 1013 AA.
XX
AC AAB26179;
XX
DT 12-FEB-2001 (first entry)
XX
```

```
DE Human CASB619 protein #1.
XX
KW Human; CASB619; cancer; autoimmune disease; immunogen; vaccine;
epitope.
XX
OS Homo sapiens.
XX
PN WO200058460-A2.
XX
PD 05-OCT-2000.
XX
PF 20-MAR-2000; 2000WO-EP02478.
XX
PR 26-MAR-1999; 99CB-0007113.
PR 25-SEP-1999; 99CB-0022858.
XX
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Brück CEM, Cassart J, Coche T, Vinals De Bassols YC;
XX
DR WPI: 2000-664923/64.
DR N-PSDB; AAA95442.
XX
PT Novel CASB619 polypeptides useful for diagnosing, and as vaccines for
prophylactic and therapeutic treatment of, cancers, particularly
ovarian and colon carcinoma, and autoimmune diseases
XX
PS Claim 4; Page 54-56; 68pp; English.
XX
CC The present sequence comprises the human CASB619 protein sequence. This
protein is thought to be specifically or over-expressed in tumour cells,
and so can be used as a target for antigen-specific immune responses
CC which can cause destruction of the tumour cell. In addition, the protein
and gene can be used in cancer diagnosis, in the treatment of autoimmune
diseases and in vaccines against cancer and autoimmune disease. The
invention provides a number of epitopes derived from the protein which
can be used as immunogens.
XX
SQ Sequence 1013 AA;
Query Match 53.7%; Score 3050.5; DB 21; Length 1013;
Best Local Similarity 54.0%; Pred. No. 6.9e-231;
Matches 543; Conservative 171; Mismatches 271; Indels -21; Gaps 11;
QY 23 RGRSPWPSPAWICWALAGCAQAAWAG---DLPSSSSRPLPPCQEKDYHFEYTECDSGSR 79
Db 15 RGRTERTRPR---LWRL-----LLWAGTAFQVGTQGTDELHACKSEHYHFEYACDSTGSR 67
QY 80 WRVATPNSAVDCSGLDPLDVRGKECTFSCASGEYLEMKNOVCSCGEGTSLGSGIKFDEW 139
Db 68 WRVAVPHTPGLCTSLDPLDVRGKECTFSCNAGFELDMQDQCKPCAGRYSLGTGIRFDEW 127
QY 140 DELPAGFSNIATFMDTVVGPDSRPGDCNNSWIPRCNYIESNRDDCTVSLIYAVHLKKS 199
Db 128 DELPHGFASLNAEMELDDSAAES-TGNCSTSKWVPRGDIASNTDECTATLMAVNLKQS 186
QY 200 GYVFVEFYVDNNIEFFEFFIONDQOQEMDITTDKWKVLTDCGEMGSHSVMLKSGTNYLYW 259
Db 187 GTVNEFYVPPSSIIFFVQNDQCP-NADDSRWKMTTEKG-WEFHVELNRGNVLYW 244
QY 260 RTTGILMGSKAVKPVLVKNITIEGVAYTSCFPCKPGTFSNPGFNCQVCPNTYSEKG 319
Db 245 RTTAFSVMTKVPKPVLVKNITIEGVAYTSCFPCKPGTYADKQSGSFCCLCPANSYKNG 304
QY 320 AKECIRCKDDSQFS--GSSECTERPPCTTKDYFOIHTPCDEEGKTOIMYKWIPEKICRED 377
Db 305 ETSCHQC--DPKRYSEKSSSCNVRPACTDKDYFTHYTTACDANGETQLMKYKAKPKICSD 363
QY 378 LTDAIRLPPSGEKKDCPCNPGFYNNSSSCHPCPPGPGTSGDKTECRPCAGTEPALGFE 437
Db 364 LEGAVKLPASGVKTHCPCNPGFFKTNNTCQPCPYGSYNGS-DCTRCPCAGTEPAVGFE 422
QY 438 YKWNVLPNKMKTSCFNVGNSKCDGMNGWEVAGDHIQSCAGGSNDYILNLHIHIFGFKPP 497
```

Db 423 YKWNTPNMTTSLVSGINFEYKGMTGWEVAGDHIYTAAGASDNDFMILTLLVPGFRPP 482
Qy 498 TS-WTGATGSELGRITTFVFETLCSADCVLYFMWDINRSTWVSWGKTKEQATTHIF 556
Db 483 OSVMADTENKEVARITTFVFETLCSVNCLELYFMVGNSTNTPVETWKGSKQSYTII 542
Qy 557 KNATFTTFAFORTNQODNRFINDMVKIYSITATNAVGVASSCRACALGSEQSSG 616
Db 543 ENTTSFTWAFORTTFHEASRKYTNDAKIIYSINTVNVNGVASCRCPCALEASDVGSSC 602
Qy 617 VPCPPGHVIEKTNCKECPDPTYLIIHQVYCKEACIPCGSGKNNQDHSVCYSCDFYH 676
Db 603 TSCPAGYIIDRSGTCHSCPNTILKAHQYGVQACVPCGPGCTKNNKTHSLCYNDCTFSR 662
Qy 677 EKENQILYDFNLASVGLMNGPSTSGTKYFHFHFNISLCGHEGKKNALCTNITDFT 736
Db 663 NTPTRTFNFYSALANTVTLAGPSFTSKGLYFHFHFTLSLCGNGKRMVCTDNVTLR 722
Qy 737 VKEIVAGSDDYTNLVCAGFQSTIIPSESKGFPRAALSSQSITLADTFIGTVETTLKNIN 796
Db 723 IPE---GEGFSKSTAYVQAVIPEVGTGKAGVSSOPVSLADRLGVTMTDLGIT 779
Qy 797 IKEDMFPVPTQIPDVHFYKSTATTSCINGRSTAVKMRNPTKSGAGVISVPSKCPAG 856
Db 780 SPAELFHLESIGIPDVIFPYSNDVTSQSCSGRSTIRVRCSPOKTVPGSLLLPGTCSDG 839
Qy 857 TDCGCTFFFLWESACPLCTEDHDEHEBEGACKRGFOETLYVWNEPKWCIKISLPEK 916
Db 840 TDCGCFNHFLEWESACPLCSVADYHAIYSSCVAGIQKTYVWREPKLCSGGISLPEQR 899
Qy 917 ATCETVDFWLVKVCAGFATVALLVALTCYFWKNKQLEYKSKLVMTNNSKECELPAD 976
Db 900 TCKTIDFWLVKVISAGTCTAILLVLTCTCYFWKNKQLEYKSKLVMTNNSKECELPAD 959
Qy 977 SCAMIEGEDNEEVVYNSKOSLGLKLSLATKEKEDHFESVLKTS 1022
Db 960 SCAMIEGEDVEDDLIFTSKSLFKGKIKSTKRTPDGFDSPVLKTS 1005
RESULT 4
ID AAB83845
XX AAB83845 standard; Protein; 1013 AA.
AC AAB83845;
DT 23-JUL-2001 (first entry)
XX
DE Amino acid sequence of a human protein expressed in tumour cells.
KW Tumour cell; Immunological disease; autoimmune disease; cancer;
XX Infection.
XX Homo sapiens.
XX
FH Key
FT Peptide
FT 1..41 Location/Qualifiers
FT /note= "signal peptide"
FT Domain
FT 42..911 /note= "extracellular domain"
FT Domain
FT 912..930 /note= "transmembrane domain"
FT Domain
FT 931..1013 /note= "transmembrane domain"
XX
PN W0200131003-A1.
XX
PD 03-MAY-2001.
XX
PF 30-OCT-2000; 2000WO-FR03032.
XX
PR 29-OCT-1999; 99PR-0013629.
XX

PA (FABR) FABRE MEDICAMENT SA PIERRE.
XX Delneste Y, Magistrelli G, Jeannin P, Bonnefoy J;
XX WPT: 2001-328651/34.
DR N-PSDB; AAF89765.
XX
PT New nucleic acid, expressed in tumours and lymphoid tissue is useful for
PT identifying agents for treating tumours and autoimmune disease
XX
PS Claim 9; Page 48-51; 85pp; French.
XX
XX The present sequence represents a human protein expressed in tumour
XX cells. The polynucleotide is useful for screening cDNA/genomic DNA banks
CC and for cloning isolated DNA; identifying mutant forms of the gene that
CC encodes a human protein, where the mutations are associated with
CC abnormal gene expression, or promoters and regulators of the gene,
CC particularly for diagnosis; for recombinant expression of the derived
CC protein; as probes and primers for detection and amplification; and
CC as antisense therapeutics. The tumour expressed protein is useful for
CC raising specific antibodies and to screen agents that modulate its
CC activity, bind to it or interact with it. These agents are potentially
CC useful for treatment or prevention of diseases associated with abnormal
CC expression/activity of the protein, particularly immunological diseases
CC (autoimmune diseases and cancer) or viral, bacterial, fungal or parasitic
CC infections.
XX
SQ Sequence 1013 AA;
Query Match 53.5%; Score 3040.5; DB 22; Length 1013;
Best Local Similarity 53.9%; Pred. No. 4.2e-230;
Matches 542; Conservative 170; Mismatches 273; Indels 21; Gaps 11;
Qy 23 RCRSPWSPAWICCCWALAGCAQAAWAG---DLPSSSRRLPPCQEKDYHFEYTECDSSGR 79
Db 15 RQRTERRIPR---LWRL-----LLWAGTAFQVQTGTGPELHACKSEYHYETACDTCGR 67
Qy 80 WRVAIPNSAVDCGILPDVVRGKECTFSCASBYLEMKNQVCKSGEGYSLGSGTKFDEW 139
Db 68 WRVAVPHTPGLCTSLPDVPKGTCEGSCFSCNAGEFLDMKQDCKPCAEGRYSLGTGRFDEW 127
Qy 140 DELPAGFSNIATFMDTVVGVCPDSRPGDCNNSSWIPRGNYIESNRDDCTVSLIYAVHLKS 199
Db 128 DELPHGFASLANMELDDSAABES-TGNCYTSSKRWPRGDIASNTDECTATLAYVNLKGS 186
Qy 200 GYVFEYQYVDNNIFEFPIQNDQCEMDTTTDKWKVLTNDGEGSHSYMLKSGTNILYW 259
Db 187 CIVNFEYYPDSLIIFEFFVQNDQCP-NADDSRWMTTEKG-WEPHSELNKGNNVLW 244
Qy 260 RTTGILMGSKAVKPVLVKNITTEGVAYTSECPKPGTFSNKPSPGNCQVCPNTHYSEK 319
Db 245 RTTAFSVWTKVPKPVILVRNIATGVAYTSECPKPGTYADKQGSFCKLCPANSYSNGK 304
Qy 320 AKECIRCKDDSOFS--GSSECTERPPCTTKDYFOHTPCDEEGKTOIMYKWTPEKICRED 377
Db 305 ETSQCOC-DPDKYSEKSGSSCNVRPACTDKDYFYHTACDANGETOIMYKWKAPKICSED 363
Qy 378 LTDAIRLPPSGEKKDCPCPNPGFYNNNGSSSCHIIPCPPTFSDGTEKCRPCPACTEPALGPE 437
Db 364 LEGAVKLPAAGVKYTHCPNCPGFEKTNSTQCPYGPYNGS-DCTRCPCAGTEPAVGPE 422
Qy 438 YKWNVLPGNMKTSCFNVGNKSCDGMNGEVAGDHIYTAAGASDNDFMILTLLVPGFRPP 497
Db 423 YKWNNTLPTNMTTSLVSGINFEYKGMTGWEVAGDHIYTAAGASDNDFMILTLLVPGFRPP 482
Qy 498 TS-WTGATGSELGRITTFVFETLCSADCVLYFMWDINRSTWVSWGKTKEQATTHIF 556
Db 483 OSVMADTENKEVARITTFVFETLCSVNCLELYFMVGNSTNTPVETWKGSKQSYTII 542
Qy 557 KNATFTTFAFORTNQODNRFINDMVKIYSITATNAVGVASSCRACALGSEQSSG 616
Db 543 ENTTSFTWAFORTTFHEASRKYTNDAKIIYSINTVNVNGVASCRCPCALEASDVGSSC 602

QY 617 VPCPPGHVIEKTNCKECPDPTYLISHOVYKGEACIPCGGSGKNNQDHSVCYDCPFYH 676
Db 603 TSCPAGYVIDRSDGCHSCPNTILKAHQPYGVQACVPCGPTKNNKTHSLCYNDCTFSR 662
QY 677 EKENGILHYDFSNISSVSGSLMNGPSFTSKGPKYHFFNISLCGHEGKMAKLTNNITDFT 736
Db 663 NTPRTFNYSALANTVTLAGGSFSTKGLKYFHHTLSLCCGNOGRKMSVCTDNVTDLR 722
QY 737 VKEIVAGSDVTVNLGVAVFCOSTIIPSESKGFRAALSSQSIIADTFIGVTVTETLNIN 796
Db 723 IPE---GESGFSKSTAVXCOAVIPEVTGYKAGVSSQPVSLADRLIGVITDMLDGI 779
QY 797 IKEDMEPVTQIPIDVHFFKYSSTATSCINGRSTAYKMRNPTKSGAGVTSVPSKCPAG 856
Db 780 SPAELHLESIGIPDVIFRYNSDVTSCSSGRSTTIRVRCSPQKTPVGSLLLPFTCSDG 839
QY 857 TCDGCTFYFLWESAACPCTEHDHFHEGACKRGFOETLYWNEPKWCIGISLPEKKL 916
Db 840 TCDGCTFYFLWESAACPCTEHDHFHEGACKRGFOETLYWNEPKWCIGISLPEKKL 916
QY 917 ATCTVDFWLKVGAGVGAFTAVLLVLTCTYFVKKNQKLEYKYSKLVMTNSKECELPAAD 976
Db 900 TICKTIDFWLKVGSAGTCTAIIITVLTCTYFVKKNQKLEYKYSKLVMTNSKECELPAAD 959
QY 977 SCATMEGEDNEEYVYSKNSLLCKLKLATKEKEDHFEFVOLKTS 1022
Db 960 SCATMEGEDNEEYVYSKNSLLCKLKLATKEKEDHFEFVOLKTS 1005
RESULT 5
AAU12190
ID AAU12190 standard; Protein; 1013 AA.
XX AAU12190;
AC AAU12190;
XX 24-OCT-2001 (first entry)
XX Human PR04985 polypeptide sequence.
DE Human secretory and transmembrane; PRO; mammalian; cancer; lung;
XX breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
KW cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
KW adipocyte; A-peptide; factor VIIA; gene therapy.
XX Homo sapiens.
XX W0200140466-A2.
XX 07-JUN-2001.
XX 01-DEC-2000; 2000WO-US32678.
XX 01-DEC-1999; 99WO-US28301.
XX 01-DEC-1999; 99WO-US28634.
XX 02-DEC-1999; 99WO-US28551.
XX 02-DEC-1999; 99WO-US28564.
XX 02-DEC-1999; 99WO-US28565.
XX 09-DEC-1999; 99US-0170262.
XX 16-DEC-1999; 99WO-US30095.
XX 20-DEC-1999; 99WO-US30911.
XX 20-DEC-1999; 99WO-US30999.
XX 30-DEC-1999; 99WO-US31243.
XX 06-JAN-2000; 2000WO-US00277.
XX 06-JAN-2000; 2000WO-US00376.
XX 11-FEB-2000; 2000WO-US03565.
XX 18-FEB-2000; 2000WO-US04341.
XX 22-FEB-2000; 2000WO-US04342.
XX 24-FEB-2000; 2000WO-US04414.
XX 24-FEB-2000; 2000WO-US04914.
XX 24-FEB-2000; 2000WO-US05004.
XX 01-MAR-2000; 2000WO-US05601.
XX 20-MAR-2000; 2000WO-US07377.
XX 21-MAR-2000; 2000WO-US07532.

30-MAR-2000; 2000WO-US08439.
17-MAY-2000; 2000WO-US13705.
22-MAY-2000; 2000WO-US14042.
30-MAY-2000; 2000WO-US14941.
02-JUN-2000; 2000WO-US15264.
10-NOV-2000; 2000WO-US30873.
(GETH) GENENTECH INC.
Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;
Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
WPI; 2001-408281/43.
DR N-PSDB; AAS21262.
XX Isolated, secretory and transmembrane PRO polypeptide used to detect
other PRO polypeptides, link bioactive molecules to cells expressing
PRO polypeptides, and detect the presence of mammalian tumours e.g.
lung, breast, prostate, cervical
Claim 12; Fig 38; 813pp; English.
CC AAU12172-AAU12446 represent novel human secretory and transmembrane
PRO polypeptides. The PRO polypeptides are useful to detect other
PRO polypeptides, to link bioactive molecules to cells expressing
PRO polypeptides, to modulate biological activities of cells expressing
PRO polypeptides, and to detect the presence of mammalian lung, colon,
breast, prostate, rectal, cervical or liver tumours by comparing PRO
polypeptide expression in a cell sample to that in a control sample.
Some of the 275 sequences are also useful to stimulate the release of
tumour necrosis factor-alpha (TNF-alpha) from human blood, the
proliferation or differentiation of chondrocytes, the proliferation or
gene expression in pericyte cells, the release of proteoglycans from
cartilage, the proliferation of inner ear utricular supporting cells or
of T-lymphocytes, the release of a cytokine from peripheral blood
monocytes (PBMCs), or the proliferation of endothelial cells. Some of
the PRO polypeptides may modulate glucose or free fatty acid uptake by
skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
to factor VIIA. The PRO polypeptides can be used in assays to identify
molecules involved in binding interactions. The polynucleotides encoding
PRO polypeptides can be used to generate probes, antisense RNA/DNA,
transgenic or knock out animals and can be used in gene therapy.
XX Sequence 1013 AA:
Query Match 53.2%; Score 3025; DB 22; Length 1013;
Best Local Similarity 54.4%; Pred. No. 7e-229;
Matches 535; Conservative 169; Mismatches 265; Indels 14; Gaps 9;
QY 46 WAG---DLPPSSSRPLPPCOEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKE 102
Db 31 WAGTAFQVTOGTGTPHACKESYHYEYACDSTGSRWRVAVPHTPGLTSLSPVKCTE 90
QY 103 CTFSACAGEYLEMKNQVCKGCEGTYSLGSGIKEDENDELPAFNSIATFMDTVVGPDS 162
Db 91 CSFSCNAGEFLDMKQDCKPCAEGRYSLGTGIREDEWDELPHGFASLANNMELDSSAAES 150
QY 163 RPDGNNSSWIPRGNYTESNRDDCTVSLIYAVHLKSKYVFEYQYVDNNIFEEFIQND 222
Db 151 -TGNCSTSKWVPRGDY TASNTDECTATLMAVNLKQSGTVNFEYIYPPDSSIIFFEVQND 209
QY 223 QCQEMDTTDDKWKLTNDNGSGSHSVMLKSGTNILYWRITGILMGSKAVKPVKNITIE 282
Db 210 QCQP-NADDSRWMTTEKG-WEFHSHVELNRGNVLYWRTAFSVWTKVPKVLVRIAIT 267
QY 283 GVAYTSECFCCKPGTFNKNKPGSCFNCQVCPRTNTYSEKAKECIRCKDDSQFS--GSSECTE 340
Db 268 GVAYTSECFCCKPGTYADKQGSFCKLCPANSYSNKGETSCHQC-DPDKYSEKSGSSCNV 326
QY 341 RPPCTTKDYFOIHTPCDEEGKTOIMYKWKIEPKICREDLTDAIRLPPSGEKKDCPCPNPGF 400
Db 327 RPACTDKDYFYTHTACDANGETQLMYKWKAPKICSELEGAVKLPASGVKTHCPCPNPGF 386

XX	AC	AA83851;	356	CDEGKTOIMYKWIEPKICREDLTDALRLPPSGEKKDCPPCNPQGFYNNSSSCHPCPPGT	415
XX	DB		301	CDANGETQIMYKWKAPKICSSEDLGAVKLPASGVKTHCPCPCNPGEFTNNSTCQCPVGP	360
XX	DB		416	FSDGTEKCRPCPCAGTEPALGFYKWNVPCNMKTSFENVGNSKCDGMNGHEVAGDHIQS	475
XX	DB		361	YSNGS-DCTRCPCAGTEPALGFYKWNVPCNMKTSFENVGNSKCDGMNGHEVAGDHIQS	419
XX	DB		476	GAGSDNDYLILNLHIFGEPPTS-MTGATGSELGRITFVFETLCSADCVLYFVMDINRK	534
XX	DB		420	AAGASDNDYFMTLVVPGFRPQPMADTENKEVARITFVFETLCSVNCELYFVGVNSR	479
XX	DB		535	STNVESWGTTKEQAYTHILFNATFTTMAFORTQODNRFRINDMVKIYSITATNA	594
XX	DB		480	TNTVETWKGSKGQSYTHIEENTTSFTWAFORTTFHEASRYTNDVAKIYSINVTNV	539
XX	DB		595	VDGVASCRACALGSESGSCVPCPGHYTEKTNQCKECPDPTYLISHOVYKEACIP	654
XX	DB		540	MNGVASYCRPCALEASDVGSCTSCPAGYITDRSGTCHSCPNTILKAHQPYGVQACVP	599
XX	DB		655	CGPGSKNNQDHSVCYSCFFHYHEKENQILHYDFSLSLSSVGSIMNGPSTSGTKYFHFFN	714
XX	DB		600	CGPGTKNNKIHSLCYNDCTFSRNTPTRTFNYSALANTVTLAGPSTSGTKYFHFFN	659
XX	DB		715	ISLCGHEGKKNALCTNNITDFTVKEIVAGSDYITNLVGFVQSTIIPSESGKRAALSS	774
XX	DB		660	LSLCGQNRKMSVCTDNTDLRIPE---GESGFSKSIATYVQAVIIPPEVTGYKAGVSS	716
XX	DB		775	QSILADTFIGVTVETTLKNIKEDMPFVPTQIPDVHFFYKSTATTSCINGRSTAVK	834
XX	DB		717	QPVSLADRLIGVTTDMTLDGITSAPAELEHLSLGPDPVIFFYRSNDVTQSCSSGRSTTIR	776
XX	DB		835	MRCNPTKSGAGVISVPSKCPAGTCDGCTFYFLWESAEACPLCTEHDHIEIGACKRGFQE	894
XX	DB		777	VRCSPQKTVPSGLLPVGCSDGTCGDNFHLWESAAACPLCSVADYRAIVSSCVAGIQK	836
XX	DB		895	TLVWNEPKWCIKIGSLPEKILATCETVDLWLVK	928
XX	DB		837	TYVWREPKLCSGGISLPEQRVTICKTIDFVLKV	870
XX	DB		RESULT 9		
XX	DB		AA848377		
XX	DB		ID	AA848377 standard; Protein; 464 AA.	
XX	DB		AC	AA848377;	
XX	DB		XX	20-APR-2001 (first entry)	
XX	DB		DE	Human SEC10 protein sequence (clone ID 1795045.0.77).	
XX	DB		DE	SECX; cytostatic; gynecological; gene therapy; screening assay; human;	
XX	DB		KW	SEC10; chromosomal mapping; forensic biology; cell proliferation; cancer;	
XX	DB		KW	cell differentiation; immune associated disorder; gestational disease.	
XX	DB		OS	Homo sapiens.	
XX	DB		XX	WO200078802-A2.	
XX	DB		PN	28-DEC-2000.	
XX	DB		PD	23-JUN-2000; 2000WO-US17328.	
XX	DB		PF	23-JUN-1999; 99US-0140584.	
XX	DB		PR	20-JUL-1999; 99US-0144722.	
XX	DB		PR	16-SEP-1999; 99US-0154520.	
XX	DB		PR	22-JUN-2000; 2000US-0604286.	
XX	DB		XX	(CURA-) CURAGEN CORP.	
XX	DB		XX	Shimkets RA, Fernandes E, Vernet C, Yang M, Boldog FL;	
XX	DB		PI	Herrmann JL;	
XX	DB		XX		

Query Match 47.6%; Score 2702.5; DB 22; Length 870;
Best Local Similarity 54.1%; Pred. No. 1.4e-203;
Matches 473; Conservative 153; Mismatches 237; Indels 11; Gaps 8;

QY 58 LPPQCKDYHFEVTECDSSGRWRVAIPNSAVDCSLPDPVRGKECTFCSCASGEYLEMKN 117
DB 5 LHACKSEYHETACDSTGSRWRVAVPTPGLTSLPDPVKGTECFSCNAGEFLDMKD 64

QY 118 QVCKGCEGYISLGKIFDEWDELPAFNSIATFMDTVVGPDSRDCNCSNWSITPRGN 177
DB 65 QCKPCAGRYSILGTGIREDEWDELPHGFASLSANMELDSDAAES-TGNCTSKWYPRGD 123

QY 178 YTESRDDCTVSLIYAVHLKSGYVFEQYVDNNTFFEFFEIONDOCEMDITTDKWKVL 237
DB 124 YIASNTDECTATLMYAVNLKSGTVNFXYPPDSSIFEFFVNDQCQP-NADDSRWMT 102

QY 238 TDNGEGSHSVMLKSGTNTLYRTTGLMGSKAVKPLVKNITIEGVAYTSECFPCPKGT 297
DB 183 TEKG-WEFHSVELNRNNVLYWTTAFVSWTKVPKPLVNRNITGVAYTSECFPCPKGT 241.

QY 298 FSNKPGSFNCQVCPNTYSEKGAKEICRCKDDQFS--GSSECTERPPCTTKYFOIHTP 355
DB 242 YADKQGSFCKLCPANSYKNGETSCHQCDPDYKSEKSGSSCNVRPACTDKDYFYTHA 300

DR WPI: 2001-071385/08.
XX N-PSDB; AAC84891.
XX Polynucleotides encoding SECX proteins useful for treating disease
PT characterized by an aberrant level of cell proliferation and/or
PT differentiation like cancer or immune associated disorders -
XX
PS Claim 1; Fig.10; 132pp; English.
XX The invention relates to human SECX polypeptides and polynucleotides
CC encoding them. The SECX polypeptides can be expressed by standard
CC recombinant methodology. The SECX polypeptides are useful for treating
CC or preventing a SECX-associated disorder. The invention is useful in
CC screening assays; detection assays (e.g. chromosomal mapping, cell and
CC tissue typing, forensic biology); predictive medicine (diagnostic assays,
CC prognostic assays, monitoring clinical trials, and pharmacogenomics); and
CC methods of treatment (e.g. therapeutic and prophylactic), especially
CC disorders characterized by aberrant cell proliferation and/or
CC differentiation like cancer or immune associated disorders or gestational
CC disease. The present sequence represents a SEC10 protein.
XX
SQ Sequence 464 AA:
Query Match 44.5%; Score 2526; DB 22; Length 464;
Best Local Similarity 98.9%; Pred. No. 4.3e-190;
Matches 452; Conservative 0; Mismatches 3; Indels 2; Gaps 1;
QY 115 MKNQVCSKCGCTGYSLGSGIKFDEWDELPAGFSNIATFMDTVVGPDSRPGDCNNSWIP 174
DB 1 MKNQVCSKCGCTGYSLGSGIKFDEWDELPAGFSNIATFMDTVVGPDSRPGDCNNSWIP 60
QY 175 RGNYESNRDDCTVSLIYAVHLKSGYVFEYQYVDNIIFFEFFIQNDQCQEMDTTIDKW 234
DB 61 RGNYESNRDDCTVSLIYAVHLKSGYVFEYQYVDNIIFFEFFIQNDQCQEMDTTIDKW 120
QY 235 VKLTONGEWSHVMKSGTNLIYWRITGILMGSKAVKPVLVKNITIEGVATSECFPCPK 294
DB 121 VKLTONGEWSHVMKSGTNLIYWRITGILMGSKAVKPVLVKNITIEGVATSECFPCPK 180
QY 295 PGTFSNKGPSFNCQVCPNRYSEKAKECIRCKDDSOFS--GSSECTERPPCTTKDYFOI 352
DB 181 PGTFSNKGPSFNCQVCPNRYSEKAKECIRCKDDSOFS--GSSECTERPPCTTKDYFOI 240
QY 353 HTPCDEEGKTQIMYKWIPEKICREDLTDAILRPPSGEKKDCPPCNPFGYNGSSSCHPCP 412
DB 241 HTPCDEEGKTQIMYKWIPEKICREDLTDAILRPPSGEKKDCPPCNPFGYNGSSSCHPCP 300
QY 413 PCTFSDGTKECRPCPAGTEPALGFYKWNVLPGNMKTSFCFNVGNSKCDGMNGWEVAGDH 472
DB 301 PCTFSDGTKECRPCPAGTEPALGFYKWNVLPGNMKTSFCFNVGNSKCDGMNGWEVAGDH 360
QY 473 IQSGAGGSDNDYLIILNLHIPGFKPPTSMGTGATGSELGRITTFVFTLCSADCVLYFMVDIN 532
DB 361 IQSGAGGSDNDYLIILNLHIPGFKPPTSMGTGATGSELGRITTFVFTLCSADCVLYFMVDIN 420
QY 533 RKSTNVVESWGCTKEKQAYTHIIFKNATFTTFAQF 569
DB 421 RKSTNVVESWGCTKEKQAYTHIIFKNATFTTFAQF 457
RESULT 10
AAB48372 ID AAB48372 standard; Protein; 411 AA.
XX
XX AAB48372;
XX
XX 20-APR-2001 (first entry)
XX Human SEC5 protein sequence (clone ID 1795045.0.61).
DE SEC5; cytostatic; gynecological; gene therapy; screening assay; human;
KW SEC5; chromosomal mapping; forensic biology; cell proliferation; cancer;
KW cell differentiation; immune associated disorder; gestational disease.

XX Homo sapiens.
XX WO200078802-A2.
XX 28-DEC-2000.
XX 23-JUN-2000; 2000WO-US17328.
XX 23-JUN-1999; 99US-0140584.
PR 20-JUL-1999; 99US-0144722.
PR 16-SEP-1999; 99US-0154520.
PR 22-JUN-2000; 2000US-0604286.
XX (CURA-) CURAGEN CORP.
XX Shinketsu RA, Fernandes E, Vernet C, Yang M, Boldog FL;
PI Herrmann JL;
XX WPI: 2001-071385/08.
DR N-PSDB; AAC84886.
XX Polynucleotides encoding SECX proteins useful for treating disease
PT characterized by an aberrant level of cell proliferation and/or
PT differentiation like cancer or immune associated disorders -
XX
PS Claim 1; Fig 6; 132pp; English.
XX The invention relates to human SECX polypeptides and polynucleotides
CC encoding them. The SECX polypeptides can be expressed by standard
CC recombinant methodology. The SECX polypeptides are useful for treating
CC or preventing a SECX-associated disorder. The invention is useful in
CC screening assays; detection assays (e.g. chromosomal mapping, cell and
CC tissue typing, forensic biology); predictive medicine (diagnostic assays,
CC prognostic assays, monitoring clinical trials, and pharmacogenomics); and
CC methods of treatment (e.g. therapeutic and prophylactic), especially
CC disorders characterized by aberrant cell proliferation and/or
CC differentiation like cancer or immune associated disorders or gestational
CC disease. The present sequence represents a SEC5 protein.
XX
SQ Sequence 411 AA:
Query Match 39.1%; Score 2222; DB 22; Length 411;
Best Local Similarity 98.5%; Pred. No. 3e-166;
Matches 397; Conservative 1; Mismatches 3; Indels 2; Gaps 1;
QY 169 NSSWIPRGNYESNRDDCTVSLIYAVHLKSGYVFEYQYVDNIIFFEFFIQNDQCQEMD 228
DB 2 NSSWIPRGNYESNRDDCTVSLIYAVHLKSGYVFEYQYVDNIIFFEFFIQNDQCQEMD 61
QY 229 TTTDKWVLTONGEWSHVMKSGTNLIYWRITGILMGSKAVKPVLVKNITIEGVATYS 288
DB 62 TTTDKWVLTONGEWSHVMKSGTNLIYWRITGILMGSKAVKPVLVKNITIEGVATYS 121
QY 289 ECFPCPKGTFSNKGPSFNCQVCPNRYSEKAKECIRCKDDSOFS--GSSECTERPPCTT 346
DB 122 ECFPCPKGTFSNKGPSFNCQVCPNRYSEKAKECIRCKDDSOFS--GSSECTERPPCTT 181
QY 347 KDYFOIHTPCDEEGKTQIMYKWIPEKICREDLTDAILRPPSGEKKDCPPCNPFGYNGSS 406
DB 182 KDYFOIHTPCDEEGKTQIMYKWIPEKICREDLTDAILRPPSGEKKDCPPCNPFGYNGSS 241
QY 407 SCHPCPPGTFSNKGPSFNCQVCPNRYSEKAKECIRCKDDSOFS--GSSECTERPPCTT 466
DB 242 SCHPCPPGTFSNKGPSFNCQVCPNRYSEKAKECIRCKDDSOFS--GSSECTERPPCTT 301
QY 467 EVAGDHIOGAGGSDNDYLIILNLHIPGFKPPTSMGTGATGSELGRITTFVFTLCSADCVLY 526
DB 302 EVAGDHIOGAGGSDNDYLIILNLHIPGFKPPTSMGTGATGSELGRITTFVFTLCSADCVLY 361
QY 527 FMVDINRKSTNVVESWGCTKEKQAYTHIIFKNATFTTFAQF 569
DB 362 FMVDINRKSTNVVESWGCTKEKQAYTHIIFKNATFTTFAQF 404

RESULT 11

AAB35328 standard; Protein: 750 AA.

AC AAB35328;

DT 08-MAY-2001 (first entry)

DE Human TR13 receptor protein SEQ ID NO: 2.

KW Human: tumour necrosis factor receptor; TR13; TR14; infection;
cancer; autoimmune disease; allergy; inflammatory disease;
graft rejection; apoptosis; cardiovascular disease; aneurysm.

OS Homo sapiens.

PN WO200105834-A1.

PD 25-JAN-2001.

PF 14-JUL-2000; 2000WO-US19343.

PR 16-JUL-1999; 99US-0144087.

PR 18-AUG-1999; 99US-0149450.

PR 20-AUG-1999; 99US-0149712.

PR 10-SEP-1999; 99US-0153089.

XX (HUMA-) HUMAN GENOME SCI INC.

PA Ruben SM, NI J, Young PE;

PI WPI: 2001-112682/12.

DR N-PSDB; AAF27997.

XX Nucleic acids encoding 2 human tumor necrosis factor receptor

PT polypeptides ((TR13) and ((TR14))), useful for the prevention, diagnosis

CC and treatment of, e.g. cancers, acquired immune deficiency syndrome and

CC hypohidrotic ectodermal dysplasia.

CC Claim 40; Page 369-372; 418pp; English.

XX The present invention provides the protein and coding sequences of the

CC human tumour necrosis factor receptors TR13 and TR14. These sequences are

CC useful in the diagnosis and treatment of many diseases, including cancer,

CC autoimmune diseases, cardiovascular disorders, allergies,

CC neurodegenerative diseases, graft rejection, inflammation, aneurysms and

CC infections.

XX Sequence 750 AA:

Query Match 38.1%; Score 2167; DB 22; Length 750;

Best Local Similarity 54.2%; Pred. No. 1.5e-161;

Matches 380; Conservative 118; Mismatches 195; Indels 8; Gaps 5;

QY 308 QVCPRTNTYSEKAKCIRCKDDSQFS--GSSECTERPPCTTKDYFQIHTPCDEGKQTQIM 365

DB 42 QTLPSNSYNKGETSCHQC-DPDKYSEKSGSSCNVRPACTDKDYFVTHACDANGETQLM 100

QY 366 YKWIEPKICREDLTDAILRPPSGEKKDCPPCNPFGYNNSSSCHPCPGTFSDCGKCECP 425

DB 101 YKWAKEKICSEDLGAVKLPAKGVKTHCPNPGFFKTNSTCQPCPYGSYNGS-DCTR 159

QY 426 CPAGTEPALGFYKKNVLPGNMKTSFNVGNSKCDGMNGWEVAGDHIOGAGGSDNDYL 485

DB 160 CPAGTEPAGVFYKKNVTLPTNMETTTLVLSINFYKGMTGWEVAGDHIYTAAGASNDPM 219

QY 486 ILNLHPIGPKPPTS-MTGAATGSELGRITVFETLCSADCVLYPMVDINRKTNVVSWG 544

DB 220 ILTLVPGFRPPQSVQADTENKEVARITFVETLCSVNCLEYPMGVGNSRNTNTPVETWK 279

QY 545 TKEQAYTHIIFKNATFTTFAWQRTNOGODNRRFINDMKVIYSITATNAVDGVASSCRA 604

Db 280 SKGQSYTYIIIEENTTSFTWAFQRTTFHEASRYKTYNDVAKIYSINVTNVMNGVASVCRP 339

QY 605 CALGSEQSGSCVPCPPGHYIEKETNOCKECPDPTYSIHQVYKEACIPGCGPKSNQD 664

Db 340 CALDASDVGSCTSCPAGYIDRDSGCTCHSCPNTILKAHQPGVQACVPCGPGTKNNKI 399

QY 665 HSCVSDCFFYHEKENQILHYDFSNLSSVGLMNGPSTSKGTKYFHFNFNLSLGHGCK 724

Db 400 HSLCYNDCTFSRNTPTRTFNYSALANTVTLAGPGPSTSKGLKYFHFHTLSLGCNQGRK 459

QY 725 MALCTNNITDFTVKEIVAGSDDYTNLVCFAVCQSTIIIPSESKGFPAAALSSQSIILADTFI 784

Db 460 MSVCTDNVTDLRIPD--GESGFSKSIYAYVQAVIIPPEVTGYKAGVSSOPVSLADRLI 516

QY 785 GVTVEITLKNINIKEDMPVPTSQIPDVHFFYKSSATTSTCINGRSTAVKVRNCPNPKSGA 844

Db 517 GVTDMTLDGITSAPLEHLSLGIPIDVIFFRSNDVTQSCSSGSRSTIRVRCSPQKTPV 576

QY 845 GVISYSPKCPAGTCDCGCTFYFLWESAECPLCTEHDHFEIEGACKRGQETLYVWNEPKW 904

Db 577 GSLLLPGTCSGTCDCGNCNPHFLWESAAACPLCSVADYHAIIVSSCVAGIQKTTYVWREPKL 636

QY 905 CIKGISLPEKKLATCETVDFWLKVGAGVCAFTAVLLVALTCYFWEKKKOKLEYKSKLYMT 964

Db 637 CSGGISLPEQRVTICKTIDFWLKVGISAGTCTPAILLTTLTCYFWEKKKOKLEYKSKLYMN 696

QY 965 TNSKECELPAADSCAIMGEDNEEEVYVSNKQSLGKIKSL 1005

Db 697 ATLKCDDLPAADSCAIMEGEDVEDDLIFTSKNIHSLGRSNHL 737

RESULT 12

AAV59972

ID AAV59972 standard; Protein: 495 AA.

XX AC AAV59972;

DT 31-JAN-2000 (first entry)

XX Human endometrium tumour EST encoded protein 32.

XX Endometrium; human; tumour; cancer; anticancer; cytostatic; EST;

KW treatment; uterine; gene therapy; expressed sequence tag.

XX OS Homo sapiens.

PN DE19817948-A1.

XX PD 21-OCT-1999.

XX PF 17-APR-1998; 98DE-1017948.

XX PR 17-APR-1998; 98DE-1017948.

XX PA (META-) METAGEN GES GENOMFORSCHUNG MBH.

XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pillarsky C, Dahl E;

XX WPI: 1999-591957/51.

XX DR N-PSDB; AA241991.

XX New nucleic acid sequences expressed in uterine cancer tissues, and
derived polypeptides, for treatment of uterine and endometrial cancer
and identification of therapeutic agents -

XX Claim 23; Page 287; 444pp; German.

XX This invention describes novel human nucleic acid (cDNA) sequences (A),
that are highly expressed in uterine tumour tissue and which have
anticancer and cytostatic activity. (A) are used (i) for recombinant
expression of polypeptides (B) and (ii) to isolate complete genes. (B)
are used (i) to identify agents suitable for treatment of uterine or

RESULT 14
AAB85768
ID: AAB85768 standard; Protein: 372 AA.
XX
AC
XX AAB85768;
XX

```
DT 29-OCT-2001 (first entry)
DE Human seven-transmembrane protein 50288 sequence.
XX
XX seven-transmembrane protein; G-protein coupled receptor; GPCR; human;
KW 17724; 50288; 31945; antiinflammatory; antitumor; cytostatic; virucide;
KW hepatotropic; immunosuppressive; gynecological; neuroprotective;
KW anti-HIV; immunostimulant; dermatological; antithrombotic; cardiac;
KW antianemic; antiparkinsonian; nephrotoxic; antithyroid; hemostatic;
KW cerebroprotective; osteopathic; analgesic; gene therapy; nootropic.
XX
OS Homo sapiens.
XX
XX WO200159117-A2.
XX
XX 16-AUG-2001.
XX
XX 12-FEB-2001; 2001WO-US04536.
XX
XX 11-FEB-2000; 2000US-0182061.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Glucksmann MA, Silos-Santiago I;
XX
XX WPI: 2001-514670/56.
XX
XX N-PSDB; AAH76195, AAH76196.
XX
XX New seven-transmembrane protein/G-protein coupled receptor polypeptides
PT and polynucleotides for diagnosing, treating seven-transmembrane
PT protein/receptor-related disorders and to identify modulators of
PT therapeutic use.
XX
XX Claim 8; Page 139-141; 144pp; English.
XX
XX The invention provides isolated seven-transmembrane protein/G-protein
CC coupled receptor polypeptides selected from 17724, 50288, 31945 proteins.
CC The polypeptides can be expressed by standard recombinant methodology.
CC Modulators of the polypeptides can be identified using a competition
CC binding assay or an assay for receptor-mediated signal transduction. The
CC polypeptides and polynucleotides are useful as reagents or targets in
CC seven-transmembrane protein/receptor assays applicable to treatment and
CC diagnosis of seven-transmembrane protein/receptor-mediated disorders
CC (see AAH76191 for a detailed description of the various disorders that
CC can be treated or diagnosed using the polypeptides). The polynucleotides
CC are useful to detect mutations in genes and gene expression products such
CC as mRNA, as antisense constructs to control gene expression and for
CC chromosome identification. The present sequence represents the human
CC seven transmembrane protein 50288 sequence.
XX
XX Sequence 372 AA:
XX
XX Query Match 15.5%; Score 883; DB 22; Length 372;
XX Best Local Similarity 51.7%; Pred. No. 6.4e-61;
XX Matches 103; Conservative 54; Mismatches 84; Indels 14; Gaps 7;
XX
QY 23 RGRSPWPAWICWALAGCAAWAG---DLPSSSSRPLPPCOEKDYHFEYTECDSSGR 79
DB 15 RGRTERIPR---LWRL-----LLWAGTAFTQVTGTGTPELHACKSEYHYEYACDSTGSR 67
QY 80 WRVAIPNSAVDCGLPDPVRGECTSCASGEYLEMKNCVSCGEGTYSLGSGIKFDEW 139
DB 68 WRVAIPHTGLCTSLPDPVKGTCTSCSAGNEFLDMKQSCPKCAEGRYSLGTGIRFDEW 127
QY 140 DELPAGFSNIATFMDTVVGPSDRPGCNSSWIPRGNTIESNRDCTVSLIYAVHLKKS 199
DB 128 DELPHGFASLANWELDDSAES--TGNCTSSKWPVRGDIASNTDCTATLAYAVNLKS 186
QY 200 GYVEFFEQYVNNIIFEFFIQNDQCEMDTTDKWYKLTLDNGEWGSHVNLKSGTNIWY 259
DB 187 GTVNEFVYPPDSIIIEFFVNDQCP--NADDSRMWKTTEKG--WEFHSVNLNRGNVLYW 244
QY 260 RTTGILMGSKAVPLVKNITIEGVAYTSECPKCFGTFSNKGPGSFNCQVCPRNTYSEKG 319
XX
```

```
DB 245 RTTAFSVWTKVPKVLVRNIALTGVAITSECFPCPKGTGYADKQSSFCFKLCPANSYSNKG 304
QY 320 AKECIRCKDDSOFSQ 334
DB 305 ETSCHQC-DPDKYSG 318
RESULT 15
AAB83852
ID AAB83852 standard; Protein: 209 AA.
XX
XX AAB83852;
XX
XX 23-JUL-2001 (first entry)
XX
XX Amino acid sequence of a human protein expressed in tumour cells.
XX
XX Tumour cell; immunological disease; autoimmune disease; cancer;
KW infection.
XX
XX Homo sapiens.
XX
XX WO200131003-A1.
XX
XX 03-MAY-2001.
XX
XX 30-OCT-2000; 2000WO-FR03032.
XX
XX 29-OCT-1999; 99FR-0013629.
XX
XX (FABR ) FABRE MEDICAMENT SA PIERRE.
XX
XX Delneste V, Magistrelli G, Jeannin P, Bonnefoy J;
XX
XX WPI: 2001-328651/34.
XX
XX N-PSDB; AAF89776.
XX
XX New nucleic acid, expressed in tumours and lymphoid tissue is useful for
PT identifying agents for treating tumours and autoimmune disease.
XX
XX Claim 10; Page 71-72; 85pp; French.
XX
XX The present sequence represents a human protein expressed in tumour
CC cells. The polynucleotide is useful for screening cDNA/genomic DNA banks
CC and for cloning isolated DNA; identifying mutant forms of the gene that
CC encodes a human protein, where the mutations are associated with
CC abnormal gene expression, or promoters and regulators of the gene,
CC particularly for diagnosis; for recombinant expression of the derived
CC protein; as probes and primers for detection and amplification; and
CC as antisense therapeutics. The tumour expressed protein is useful for
CC raising specific antibodies and to screen agents that modulate its
CC activity, bind to it or interact with it. These agents are potentially
CC useful for treatment or prevention of diseases associated with abnormal
CC expression/activity of the protein, particularly immunological diseases
CC (autoimmune diseases and cancer) or viral, bacterial, fungal or parasitic
CC infections.
XX
XX Sequence 209 AA:
XX
XX Query Match 12.4%; Score 705.5; DB 22; Length 209;
XX Best Local Similarity 56.2%; Pred. No. 2.6e-47;
XX Matches 126; Conservative 26; Mismatches 55; Indels 17; Gaps 3;
XX
QY 310 CPRNTYSEKGAKECIRCKDDSOFSGSSSECTERPPCTTKDYFQIHTPCDEGKTQIMYKWI 369
DB 2 CDPDKYSEK-----GSSSCNVPRACCTDKDYFYTHTACDANGETQLWYKWA 46
QY 370 EPKICREDLTDALRLPPSGEKKDCPPCNPFYNNNGSSSCHPCPPGTFSDGTEKRCPCAG 429
DB 47 KPKICSEDLEGAVKLPASGVKTHICPPCNPFYNNNGSSSCHPCPPGTFSDGTEKRCPCAG 105
QY 430 TEPALGFYKWNWLVLCNMTKSCFNVGNSKCDGMNGWEVAGDHIQSGAGGSDNDYLNL 489
XX
```

Db 106 TEPAVGEYKWNLTPTNMTETVLSGINFYKGMTGWEVAGDHIYTAAGASDNDPMILT 165
Qy 490 HIPGFKPPTS-MTGATGSELGRITVFETLCSADCVCVLYFMVDIN 532
Db 166 VVGFRRPPOSVMADTENKEVARITVFETLCSVNCCELYFMVGN 209

Search completed: May 12, 2003, 13:18:55
Job time : 53.5437 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 13:07:53 ; Search time 22.1915 Seconds
(without alignments)
1361.665 Million cell updates/sec

Title: US-10-073-333A-4
Perfect score: 5681
Sequence: 1 MLFRAGPVRGRGWGPAEA.....KEKEDHFESVOLKTSRSPNI 1027

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	165	2.9	3111	US-08-460-309-4	Sequence 4, Appli
2	165	2.9	3111	US-08-125-077-4	Sequence 4, Appli
3	162	2.9	1193	US-08-400-159-10	Sequence 10, Appl
4	162	2.9	1193	US-08-611-729A-10	Sequence 10, Appl
5	162	2.9	1253	US-08-479-722B-4	Sequence 4, Appli
6	157.5	2.8	1219	US-08-882-046-5	Sequence 5, Appli
7	155.5	2.7	1251	PCT-US95-02251-3	Sequence 3, Appli
8	155.5	2.7	1252	US-08-199-780-3	Sequence 3, Appli
9	155.5	2.7	1252	US-08-316-650-3	Sequence 3, Appli
10	153	2.7	1940	US-08-644-271-30	Sequence 30, Appl
11	153	2.7	1940	US-09-077-955-34	Sequence 34, Appl
12	151	2.7	2594	US-08-718-388-7	Sequence 7, Appli
13	151	2.7	5405	US-08-718-388-9	Sequence 9, Appli
14	149.5	2.6	2523	US-08-185-432-18	Sequence 18, Appl
15	149.5	2.6	2523	US-08-899-232-3	Sequence 3, Appli
16	149	2.6	610	US-08-365-470-3	Sequence 3, Appli
17	149	2.6	610	US-09-209-668-19	Sequence 19, Appl
18	149	2.6	610	US-09-009-490A-89	Sequence 89, Appl
19	148	2.6	2703	US-08-183-432-19	Sequence 19, Appl
20	148	2.6	2703	US-08-899-232-4	Sequence 4, Appli
21	145	2.6	610	521870-2	Patent No. 521870
22	143	2.5	1404	US-08-400-159-2	Sequence 2, Appli
23	143	2.5	1404	US-08-611-729A-2	Sequence 2, Appli
24	137.5	2.4	833	US-09-013-895A-5	Sequence 5, Appli
25	137.5	2.4	833	US-09-448-868-5	Sequence 5, Appli
26	137.5	2.4	2471	US-08-185-432-16	Sequence 16, Appl
27	137.5	2.4	2471	US-08-083-590A-19	Sequence 19, Appl

28 137.5 2.4 2471 3 US-08-532-384-19 Sequence 19, Appli
29 137.5 2.4 2471 4 US-08-899-232-1 Sequence 1, Appli
30 137 2.4 1010 4 US-08-882-046-7 Sequence 7, Appli
31 137 2.4 1036 4 US-09-068-740A-6 Sequence 6, Appli
32 137 2.4 1187 4 US-09-068-740A-7 Sequence 7, Appli
33 137 2.4 1208 4 US-09-199-865-1 Sequence 1, Appli
34 137 2.4 1218 3 US-08-400-159-6 Sequence 6, Appli
35 137 2.4 1218 3 US-08-611-729A-6 Sequence 6, Appli
36 137 2.4 1218 4 US-08-882-046-2 Sequence 2, Appli
37 137 2.4 1218 4 US-09-214-278-7 Sequence 7, Appli
38 137 2.4 1218 4 US-09-068-740A-11 Sequence 11, Appli
39 137 2.4 1218 4 US-09-467-997-1 Sequence 1, Appli
40 136.5 2.4 1276 4 US-08-937-236-3 Sequence 3, Appli
41 136.5 2.4 1291 4 US-08-569-214-3 Sequence 3, Appli
42 136.5 2.4 1291 4 US-08-937-236-2 Sequence 2, Appli
43 136.5 2.4 1295 4 US-08-569-214-2 Sequence 2, Appli
44 135.5 2.4 810 2 US-08-820-170A-34 Sequence 34, Appli
45 135.5 2.4 810 3 US-09-055-699-34 Sequence 34, Appli

ALIGNMENTS

RESULT 1
US-08-460-309-4
; Sequence 4, Application US/08460309
; Patent No. 5837496
; GENERAL INFORMATION:
; APPLICANT: Engvall, Eva
; APPLICANT: Leivo, Ilmo
; TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
; TITLE OF INVENTION: Fragments and Uses Thereof
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,309
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/125,077
; FILING DATE: 22-SEP-1993
; APPLICATION NUMBER: US PCT/US 94/10730
; FILING DATE: 21-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/472,319
; FILING DATE: 30-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/919,951
; FILING DATE: 27-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LA 9721
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3111 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-460-309-4

Query Match 2.9%; Score 165; DB 2; Length 3111;
Best Local Similarity 19.3%; Pred. No. 3.2e-05;
Matches 173; Conservative 83; Mismatches 308; Indels 332; Gaps 50;

Qy 290 CFPCKPGTFSNKP---GSF---NCQVCPNTYSEK-----GAK 321
Db 318 CQCCPG-FHQKPRAGTFLTKTEACNCHGKAECCYDENVARNLSLIRGKYIGG 376
Qy 322 ECIRCKDDSQSGSSECTE---RPPCTTKDYFOIHTP---CDEEGTKQIMYKWKIEPKICRE 376
Db 377 VCINCTQNTAGINCETCTDGFPRKGVSPNPRCPQCHCDPIGSLN-----EVCVK 428
Qy 377 DLTDAIR-LPPSGEKDCCPPNPGYNNSSSCHPCPTGTSKGKRCPC-----PAG 429
Db 429 DEKHARRGLAPG---SC-HCKTGF---GGVSCDRCARG---YTGYPDCRACNCSGLGSKN 478
Qy 430 TEPALG-----PEYKWNVLPGNMK---TSCFNVNGS-----KCD 461
Db 479 EDCFCGPICKENVEGGDCSRCKSGFFNLQEDNWKGCDECFCSGVSNRQSSWYTWGKIQ 538
Qy 462 GMNGNEVAG---DHIOGAGGSDND---YLILNLHIJPGFKPPTSMGA---TGSEL--- 508
Db 539 DMSGYLTDLPGRIRVAPQDDLDSPQOISISNAEARQALPHSYWSAPAPYLGNKLPAV 598
Qy 509 -GRITVF-----ETLCSADCVLYFMVDINRSTNVESWGCTKEQAYTH-----I 554
Db 599 GGLTFTISYDLEEEEDTERVLQMLILEGNDLSI-----STAQDEVYLHPSEETNVL 653
Qy 555 IFKNATFT-----FTWAFORTNOQDNRFRINDVMKIYS 588
Db 654 LKKEESTHTGHPVRKKEFTVLANKRVLQITYSF-----GMD-----AIFRUS 702
Qy 589 ITATNAV-----DG-VASSCRACALGSESGSCVPCPPCH----- 623
Db 703 VLESASVPTDGTAAAVECOPPGYTGSCSCSWPHRRVNGTIFGGICEPCQCFCH 762
Qy 624 --YIEKETNQCCEPPDYLYSHQVYKACIPCGP-----SKNNDHVSVCYSDCFY 675
Db 763 AESCDVDTGECLNCKDHT-----GGPYCDKCLPGFYGEPTKGTSEDPCPCAPLNIP 814
Qy 676 HEKENQILHYDFSNLSSVGLMNGPSFTSGTKYHFFFNISLC--GHEGKMKALCT--- 729
Db 815 SNFSPCTCHD-----RSLGLICDG-----CPVGYTCPRCAEGYF 852
Qy 730 -----NNITDFTVKEIVAGSDDYTNLVGA-FVCQSTIIPSESKGFRAALS 773
Db 853 GQPSVPGGSCQPCQNDLDFS-----IPGSCD--SLSGSLICKP-----GTTG 895
Qy 774 SOSIILADTFIGVYVETTLKNIKEDMFPVPTSQIPDVHFFYKSTATTSCINGRSTAV 833
Db 896 RYCELCADGYFGDAVDA--KNQ-----PCRCNAGGSFSEVCHSQTQCEC---RANVQ 944
Qy 834 KMRNPTKSGAGVISVPSKCPACTCGCTFYFLWESAEACPLCTEHDHFEIEGACKRGFO 893
Db 945 GORCD-----KCKAGT-----FGLQARCCVPCNCSFGSKSFDCEESQ 984
Qy 894 ETLVYVNEPKWCKIGSLPEKPLATCETVDFWLKVGAGVGAFTAVLLVALTCYFWKKN-- 951
Db 985 -----CWQPGYT--GKKCDRCAHGYFNFQEGG-----CTACECSHLGNND 1024
Qy 952 -----QKLEYKYSLV-----MTNSKECELPAAFS-----CAIMGEDN 986
Db 1025 PKTGRCICPPNTIGEKSCAPNTWGHSTITGCKACNCSTVGSLDFQCNVNTQCQ 1080

RESULT 2
US-08-125-077-4
Sequence 4, Application US/08125077
Patent No. 5872231
Patent No. 5872231
GENERAL INFORMATION:
APPLICANT: Engvall, Eva

APPLICANT: Leivo, Ilmo
TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/125,077
FILING DATE: 22-SEP-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US PCT/US 94/10730
FILING DATE: 21-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/472,319
FILING DATE: 30-JAN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/919,951
FILING DATE: 27-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 9721
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3111 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-125-077-4

Query Match 2.9%; Score 165; DB 2; Length 3111;
Best Local Similarity 19.3%; Pred. No. 3.2e-05;
Matches 173; Conservative 83; Mismatches 308; Indels 332; Gaps 50;

Qy 290 CFPCKPGTFSNKP---GSF---NCQVCPNTYSEK-----GAK 321
Db 318 CQCCPG-FHQKPRAGTFLTKTEACNCHGKAECCYDENVARNLSLIRGKYIGG 376
Qy 322 ECIRCKDDSQSGSSECTE---RPPCTTKDYFOIHTP---CDEEGTKQIMYKWKIEPKICRE 376
Db 377 VCINCTQNTAGINCETCTDGFPRKGVSPNPRCPQCHCDPIGSLN-----EVCVK 428
Qy 377 DLTDAIR-LPPSGEKDCCPPNPGYNNSSSCHPCPTGTSKGKRCPC-----PAG 429
Db 429 DEKHARRGLAPG---SC-HCKTGF---GGVSCDRCARG---YTGYPDCRACNCSGLGSKN 478
Qy 430 TEPALG-----PEYKWNVLPGNMK---TSCFNVNGS-----KCD 461
Db 479 EDCFCGPICKENVEGGDCSRCKSGFFNLQEDNWKGCDECFCSGVSNRQSSWYTWGKIQ 538
Qy 462 GMNGNEVAG---DHIOGAGGSDND---YLILNLHIJPGFKPPTSMGA---TGSEL--- 508
Db 539 DMSGYLTDLPGRIRVAPQDDLDSPQOISISNAEARQALPHSYWSAPAPYLGNKLPAV 598
Qy 509 -GRITVF-----ETLCSADCVLYFMVDINRSTNVESWGCTKEQAYTH-----I 554
Db 599 GGLTFTISYDLEEEEDTERVLQMLILEGNDLSI-----STAQDEVYLHPSEETNVL 653
Qy 555 IFKNATFT-----FTWAFORTNOQDNRFRINDVMKIYS 588

```
Db 654 LLKESFTIHTHFPVRKPFMTVLANKRVLLOITYSF-----GMD-----AIFRLSS 702
Qy 589 ITATNAV-----DG-VASSCRACALGSESGSCVPCPPGH-----623
Db 703 VNLESASVPTDGSAAAVEVCQPPGYTGSRCSCWPHRRVNGTIFGGICEPCQCEGH 762
Qy 624 --YIEKTNQCKECPDPYLSIHQVYGEACIPCGP-----SKNNDHVSVCYSDCFY 675
Db 763 AESCDDVTGECLNCKDHT-----GGPYCDKCLPGFYGEPTKGTSEDQPCACPLNIP 814
Qy 676 HEKENQILHYDFSLSVSGSLMNGPSFTSKTKYFHFNLSLC--GHEGKMAICT-----729
Db 815 SNNFSPTCHLD-----RSLGLICDG-----CPVGYTGPCECAEAGYF 852
Qy 730 -----NNITDFTVKEIVAGSDDYTNLGA-FVCOSTIIPSESKGFRAALS 773
Db 853 GOPSVPGSCQPCQNDNLDFS-----IPGSCD--SLSGCLICRP-----GTTG 895
Qy 774 SOSILADFTTGIVVETTLKNIKINIKEDMPVPTQIPDVHFFYKSSATTSCINGRSTAV 833
Db 896 RYCELCAADYFGDAVDA--KNCQ-----PCRCNAGGSFSEVCHSQTGQCEC---RANVQ 944
Qy 834 KMRNPTKSGAGVLSVPSKPCAGTCGCTFFFLWESAEACPLCTEHDHETHEGACKRGFQ 893
Db 945 GORCD-----KCKAGT-----FGLQSARGCVPCNCNSFGSKSFDCESGQ 984
Qy 894 ETLVWNEPKWCKIGLSUPEKKLATCETVDFWLKVGAGVGAFTAVLLVALTCYFWKKN-- 951
Db 985 -----CHCQPGVT--GKKDRCAGHYFNFGEG-----CPACESHGNNCD 1024
Qy 952 -----OKLEYKYSKLV-----MTTNSKECELPAADS-----CAIMEGEDN 986
Db 1025 PKTGRICPPNTIGEKSKCAPNTWGH5ITTGCKRACNCSTVGSLLDFQCNVNTGQCN 1080

RESULT 3
US-08-400-159-10
: Sequence 10, Application US/08400159
: Patent No. 5869282
: GENERAL INFORMATION:
: APPLICANT: Ish-Horowitz, David
: APPLICANT: Henrique, Domingos M.P.
: APPLICANT: Lewis, Julian H.
: APPLICANT: Myat, Anna M.
: APPLICANT: Fleming, Robert J.
: APPLICANT: Artavanis-Tsakonas, Spyridon
: APPLICANT: Mann, Robert S.
: APPLICANT: Gray, Grace E.
: TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE
: TITLE OF INVENTION: SERRATE GENE AND METHODS BASED THEREON
: NUMBER OF SEQUENCES: 20
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Pennie & Edmonds
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10036-2711
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/400,159
: FILING DATE: 07-MAR-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Mistrock, S. Leslie
: REGISTRATION NUMBER: 18,872
: REFERENCE/DOCKET NUMBER: 7326-029
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 790-9090
```

```
: TELEFAX: (212) 869-9741/8864
: TELEX: 66141 PENNIE
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1193 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-400-159-10
```

```
Query Match 2.9%; Score 162; DB 2; Length 1193;
Best Local Similarity 18.6%; Pred. NO. 1.3e-05;
Matches 237; Conservative 116; Mismatches 420; Indels 498; Gaps 65;
```

```
Qy 91 CSGLPDPVGRGECTSCASGEVLEMKNOVCSCGEGTYSLSGSIKGFDEHDELPAGFSNTA 150
Db 29 CDGTRNP--GDKRKT-----RDECDTYFKVCLKEYOSRVTAGG-----PCSFSGSKS 72
Qy 151 TFMDDTVVGPSSDRPDGNCNNSSWIPRGNYIESNRDCTVSLIYAVHLKKSQYVF--FEYQV 209
Db 73 T---PVIG-----GN-----TFNLKYSRNNKRNRIPIPTFAWP 103
Qy 210 DN-NITFFEFFIQNDQCEQMDTTDKWVLTDNGENGSHSVMLKSGTNTILYWRT---TGI 264
Db 104 RSYTLLEAVEADYNDNSTNPDRILIEK-----ASHSGMINPSRQ---WQTLKHNTCA 150
Qy 265 LMGSKAVPVLVKNTITIEGVAITSECFP--CKPGTFSNKPFGSPFCQVCPNRTYSEKGAREC 323
Db 151 AHFEYQIR-----VTCAEHYVFGCNCNKFGRP-----RDDFFTHHTCDQN-----GNKTC 194
Qy 324 IRCKDDSOFSGSSECTE---RPPCTTKDYFIHTPCDEEGKTOIMYKW-----368
Db 195 L-----EGWTG--PECNKAICROGCSPK-----HGSCVPGECRCQYWGQGVCDKCIPIHP 243
Qy 369 -----IEP-----KICREDLTDAIRLPP-----SGEKKDCPPCNPQGY 401
Db 244 GCVHGTCTIEPWQCLCETNNGGOLCDKDLNYCGTHPCLNGGTCSTNTPDKYQCSCEPGYS 303
Qy 402 NNG-----SSSCH-----PCPPG-----TFSDGTKECRPCPAG-----429
Db 304 GQNCETABHACLSDPCNNGSGSLETSTGFEVCAPGAGWAGTCTDNDIDCSPNCPGCGGTC 363
Qy 430 TEPALGFEY---KWNVNL-----PGNMKTSCFN--VGNKSCDCGMNWE-----467
Db 364 QDLVDGFKCICPPQWTGKTCLQDLANECEGKPCVNVANSRNLIGSYCYDCITGWSGHNCDI 423
Qy 468 -----VAGDHQSGA-----GSDNDYL---485
Db 424 NINDRCGQCGGSCRDLVNGYRCISPGYAGDHCEKIDINECASNPMMNGGHCQDEINGF 483
Qy 486 -----ILLNLHIPGFKPPTSMGTATGSELGRITFV-----FE-----516
Db 484 QCLCPAGFSGNLCQLDIDYCEPNPCQNGAQCFNLAMDYFCNCPEDYEGKNCNHLKHCHRT 543
Qy 517 TLCSA--DCVLYFMVD-----INRKSTNVVSWGTGKTKKQAYTHIIFKNATFTTFAWFR 569
Db 544 TPCEVIDSDCTVAVASNSTPEGVRISSNVCGPHGKCKSQAGGKFTCECNKGTGTGYCHEN 603
Qy 570 TNOGQDNRRFINDMVKIYISITATNAVGVASSCRACALGSESGSCVPCPPGHIYERET 629
Db 604 INDCESNP-----CKNGGTGIDGVNSYKICISDGWE-----GTYCETNI 642
Qy 630 NOCKECPDPTVLSIHQVYGEACIPCGP-----SKNNO--DHSVCYSDCFYIEKENQ 681
Db 643 NDCSKNPNCHNGGTCRDLDVNDFFC--ECKNGWKGKTCNHSRDSQCDDEATCNGGTCYDEGDT- 700
Qy 682 ILHYDFSNLSSVYG-----SLMNGPSTFSKTKYFHEFNISL 717
Db 701 -----FKCMCPAGWEGATCNLARNSSCLPNCHNGGTCVWSGDSFT-----CV 743
Qy 718 CGHEGKMAICTNNTITDFTVKEI-----VAGSDDYTNLVAFAVQ 757
Db 744 C-KEGWEGTCTQNTNDCSPHPCYNSGTCTVDGDNWYRCECAPGAGPDCRINI---NECQ 799
```

Qy 758 ST-----IIPSEKGF-----RAALSSQSIILADTFIGVTVETTLKLNINIKEDMPV 804
Db 800 SSPCAFGATCDEINGYRCICPPCRSGPGCQEVTRPCFTSIRV-----843
Qy 805 PTSQIPDVHFYKSTATTSCINGRSTAVKMRNP-----TKSGAGVISVPSK-- 852
Db 844 ---MPD-GAKWDDDCNTCQCLNGKVTCSKVMCGPRPCIIHAKGHNECPAGHACVPVKED 898
Qy 853 -----CPA-GTC-----DGC---TFYFLWESABACPLCTEHDHPE 883
Db 899 HCFTHPCAAGCEWPSNOQPKTKCNSDSYQDNCANITFTFNKEM-APGLTEHICSE 957
Qy 884 IE-GACKRGQOE-TLYVWNEPKWCI-----KGISLPEKK 915
Db 958 LRNLNILKNSAESIYITCEPSHLANNEIHVAISAEIDEDENPIKEITDKIIDLVSKR 1017
Qy 916 -----LATCET-----VDFWLKVGAGVCAFTAVLLVALTC-----YFW- 949
Db 1018 DGNNTLIAAFAEVRVQRPRVKNKDFL-----VPLSSVLTVAMICCLTVFYWCIOKR 1071
Qy 950 KQKLEYKYSVLMTTNSKECELPAAADSCAIMEGEDNEEVVYSNKQSLGLKLSLATKE 1009
Db 1072 RKQSHHTASDDNTNNVRLQNLQIKNPIEKHGANTVPIKDYENKNSKIAKIRTHNSEV 1131
Qy 1010 KEDHIFESVOLK 1020
Db 1132 BEDDMKKHQK 1142

RESULT 4

US-08-611-729A-10
; Sequence 10, Application us/08611729A
; Patent No. 6004924
; GENERAL INFORMATION:
; APPLICANT: Ish-Horowitz, David
; APPLICANT: Henrique, Domingos M.P.
; APPLICANT: Lewis, Julian H.
; APPLICANT: Myat, Anna M.
; APPLICANT: Fleming, Robert J.
; APPLICANT: Artavanis-Tsakonas, Spyridon
; APPLICANT: Mann, Robert S.
; APPLICANT: Gray, Grace E.
; TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE
; TITLE OF INVENTION: SERRATE GENE AND METHODS BASED THEREON
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/611,729A
; FILING DATE: 06-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-037
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1193 amino acids

; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-611-729A-10

Query Match 2.98; Score 162; DB 3; Length 1193;

Best Local Similarity 18.6%; Pred. No. 1.3e-05;
Matches 237; Conservative 116; Mismatches 420; Indels 498; Gaps 65;

Qy 91 CSGLPDPVRGKCTFSCASGEYLEMKNOVCKSGEGTYSLAGSIGKFEWDELPAFGSNIA 150
Db 29 CDGRNP-GDKKCT-----RDECDTYFKVCLKEYQSRVTAGG-----PCSEFGSKS 72
Qy 151 TMDTVGSPDSRDPDGCNNNSWIPRGNIENRDCCTVSLIYAVHLKSKGVYF-PEYQYV 209
Db 73 T---PVIG-----GN-----TFLKYSRNNKRNRIPIPTFAMP 103
Qy 210 DN-NIFFEFTIONDOCEMDTTTDKWKVLTNDGEMGSHSVMLKSGTNILYWR-----TGI 264
Db 104 RSYTLLEAWDYNDNSTNPDRIIEK-----ASHGMINPSRQ---WOTLKINTGA 150
Qy 265 LMGSKAVPVLVKNTIECVAYTSECFP-CKPGTFESNKPFGFNCQVCPRTIYSEKAKEC 323
Db 151 AHFEYQIR-----VTCAEHYVFGCCKEFCRP-----RDDFTHTTCQDN-----GNKTC 194
Qy 324 IRCKDDSFSGSSECTE---RPPCTTKDYFQIHPTCDEEGKTQIMYKW----- 368
Db 195 L-----EGWTG-PECNKAICRQGCSPK-----HGCTVPGECRCQYGMQGYCDKCIPIH 243
Qy 369 -----TEP-----KICREDLTDAIRLPP-----SGEKKDCPPCNPNGFY 401
Db 244 GCVHGTCTEPWQCLCETNMGGLCDKDLNYCGTHIIPCLNGGTCSMTGPDKYOCSEPGYS 303
Qy 402 NNG-----SSSCH-----PCPPG-----TFSDGTKECRPCPAG----- 429
Db 304 GQNCETAEIHAICLSDPCNHNGGSCLETSTGFECAVCPAGWAGTCTDNDICSPNPGCHGGTC 363
Qy 430 TEPALGFY-----KWNVL-----PGNMTKTSFN-VGNSKCDGMNGWE----- 467
Db 364 QDLVDGFKCICPPQWTKTQLDANECEGKPCVANSACRNLIIGSYCYDCITGWSGHNCDI 423
Qy 468 -----VAGDHQISGA-----GGSDNDYL----- 485
Db 424 NINDRCGQNGGSCRDLVNGYRCICSPYAGDICEKEDINECASNPNGGHCQDEINGF 483
Qy 486 -----ILNLHIPKFPPTSMTCATGSELGRITFV-----FE----- 516
Db 484 QCLCPAGFSNLCQLDIDYCEPNPCONGAQCFENLAMYFCNPEDYEGKNCSILKDHCHT 543
Qy 517 TLCSA--DCVLVYFMYD-----INRKSTWVSWSGTKEQATHTLIIFKNATFTTFAQR 569
Db 544 TPCEVIDSCTVAVASNSTPEGVRYISSNVCGPHGKCKSQAGGKFTCECNKGFTGTYYCHEN 603
Qy 570 TNOGODNRRFINDMVKIYSITATNAVDGVASSCRACALGSESGSSCVPCPPGHVIEKFT 629
Db 604 INDCESNP-----CKNGGTCTDGVSNYKICISGWE-----GTTCEINI 642
Qy 630 NOCKECPDPTLYLSIHQVYCKEACIPCGPG-----SKNNO-DHSVCYSDCFFYHEKENQ 681
Db 643 NDCSKNPGCHNGTCHRDLYNDFFC-ECKNGWKGTCHSRDSQCDATCNGGTCYDEGDT- 700
Qy 682 ILHYDFSNLSSVG-----SLMNGPSTTSKGTKYHEFENISL 717
Db 701 -----FKCMCPAGWEGATCNIARNSSCLPNPCHNGGTCTVWSGDSFT-----CV 743
Qy 718 CGHECKKMAICTNNITDFTVKEI-----VAGSDDYTNLVGAFVCO 757
Db 744 C-KEGEGPTCIQNTNDCSPPHCYNVSGTCVCDGNWYRCACPGFAGPCRINI---NECQ 799
Qy 758 ST-----IIPSEKGF-----RAALSSQSIILADTFIGVTVETTLKLNINIKEDMPV 804
Db 800 SSPCAFGATCDEINGYRCICPPCRSGPGCQEVTRPCFTSIRV-----843

QY 805 PTOIPDVHFFYKSTATTSCINGRSTAVKMRCP-----TKSGAGVISVPSK-- 852
Db 844 ---MPD-GAKWDDCNCQCLNGKVTCSKVMCGPRCIIHAKGHNECPAGHACVPVKED 898
QY 853 -----CPA-GTC-----DGC-----TYFFLWESAECPLCTEHDHFE 883
Db 899 HCFTHPCAAGECWPSNOQPVKTKCNSDSYODNCANITFTFNKEM-APGLTTEHCSE 957
QY 884 IE--GACKRGFQE--TLVWNEPKWCI-----KGISLPEKK 915
Db 958 LRLNLIKNSVAEYSIITCEPSHLANNEITHVAISAEDIGEDENPIKEIDKIIDLVSKR 1017
QY 916 -----LATCET-----VDFMLKVGAGVGAFTAVLLVALTC-----YFW-----K 949
Db 1018 DGNNTLIAAAVAEVRQRPVKNKTDEL-----VPLLSSVLTVAWICCLVTVEVWCICQR 1071
QY 950 KNOKLEYKSKLVMTNNSKECELPAADESCAIMEGEONEEEVYVSNKOSLLGKLSLATKE 1009
Db 1072 RKOSSHHTASDDNTNNVREQLNQIKPIEKHGANTVPIKDYENKNSKIARTHNSEV 1131
QY 1010 KEDHFSVOLK 1020
Db 1132 EEDDMKHOOK 1142

RESULT 5

US-08-479-722B-4
; Sequence 4, Application US/08479722B
; Patent No. 6074840
; GENERAL INFORMATION:
; APPLICANT: Bonadio, Jeffrey
; APPLICANT: Yin, Wushan
; TITLE OF INVENTION: LATENT TGF BINDING PROTEIN (LTBP)
; TITLE OF INVENTION: GENES, COMPOSITIONS AND METHODS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Williams, Morgan & Amerson
; STREET: 7676 Hillmont, Suite 250
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77040
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,722B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US95/02251
; FILING DATE: 21-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/316,650
; FILING DATE: 30-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/199,780
; FILING DATE: 18-FEB-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fussey, Shelley P.M.
; REGISTRATION NUMBER: 39,458
; REFERENCE/DOCKET NUMBER: 4100.000500/FUS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (713) 934-7000
; TELEFAX: (713) 934-7011
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1253 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-479-722B-4

Query Match 2.9%; Score 162; DB 3; Length 1253;
Best Local Similarity 20.3%; Pred. No. 1.4e-05;
Matches 150; Conservative 67; Mismatches 246; Indels 276; Gaps 48;

QY 289 ECFPCPKP--CTFSNKGSPNCQVCPR--NTYSEKGAKEICRCKDDSQFSSGSSCTERPPC 344
Db 601 EAEPCGPGKGCIMTGGSYNCH--CNRGYRLHVGAGGSCV-----DLNECAKPHLC 650
QY 345 TTKDY---FOIHTPCDBEGKTOIMY-----KWIEPKICREDLTDAIRLP---PSGEKKDC 393
Db 651 GDGGFCINFPGHYKCN-----CYPGYRLKASRPPIK-EDI--DECRDPSTCDPGCKENK 701
QY 394 PCNPGFYNNSSSCHIPPGTFSDTGTECR---PCPAGTEPALGFEYKWMNVLPGNMKT 450
Db 702 P-----GSPKCIACOPGYRSOGGACRDVNECSEGTPCSPG---WCENLPGSYRC 748
QY 451 SC-----FNVGNSKC---DGMNGMEVAGDHIOGAGSDNDYLILNLHIDPGFKPPTSMTG 502
Db 749 TCAQGITRTGRLSCIDVDECEAGKVCQDQIGICTNTGSCFQCCLSGYHL----- 797
QY 503 ATGSELGRITFV---FETLC--SADCVLYFMVDINRKNSTNVESW-----GGTK 546
Db 798 --SRDSRCEIDECDFPAACIGDCI-----NTNGSYRCLCPLGLHRLVGGRK 843
QY 547 -----EKOAYTHIIFKNATFTT---MAFORTNOGDNR--FIN- 581
Db 844 CKKIDECSDQGLCLPHACENLQGSYVVCDEGFTLTQDHGCEEVEQPHHKKCYLNF 903
QY 582 -DMVKIYSITATNAVGVASSCRACALGSEQSGSSC--VPC-----PPGHYIE 626
Db 904 DDTVFCDSVLATNVTQ---QEC--CCSLGAGW-GDHCEIYPCPVYSSAEFHSPLVDPGKRLH 958
QY 627 KETNOCKECPDPTYLSIHQ--VYGREACI--PCGPGSKNNQDHSVCYSCFFYHEKENOI 682
Db 959 SGQOHCELCIP-AHRDIDECILFGAETCKEGKC---VNTQPGYECYCKQGFY----- 1006
QY 683 LHYDFSNLSSVGLMNGPSTSKGTKYFFIFFNLSLCGHEGKKMALCTNNTITFTVKEIVA 742
Db 1007 --YDGNLLECV-----DVDECLDES-----CRNGVCE----- 1032
QY 743 GSDDYTNLVGAFVQOSTIIPSESKGFRAALSQSIIADTFIGVTVTETTLKNIKEDMF 802
Db 1033 -----NTRGGYRCAC-PPAEYSQAQ---QCLI-----PERM 1061
QY 803 PVPTSOIPDVHFFYKSTATTSCING-----RSTAVKMRCP--TKSGAGVI 847
Db 1062 STPQRDVKCA---GASEERTACVMPGNWAGPALTFDCCCRQPRLTQCRPCPRGTG-- 1115
QY 848 SVPSKCPAGTCDCCTFYFLWESAECAPLC-----TEHDFHEIEGACKRGFOETLYVWNE 901
Db 1116 ---SOCPTSQSESNF---WDTS---PLLLGKSPRDESDSESDSEDCR----- 1154
QY 902 PKWCIGISLPEKKLATCE 920
Db 1155 ---CVSGRCVPRPGGAVCE 1170

RESULT 6

US-08-882-046-5
; Sequence 5, Application US/08882046
; Patent No. 6136952
; GENERAL INFORMATION:
; APPLICANT: Li, Linheng
; APPLICANT: Hood, Leroy
; APPLICANT: Krantz, Ian D.
; APPLICANT: Spinner, Nancy B.
; TITLE OF INVENTION: Human Jagged Polypeptide, Encoding
; TITLE OF INVENTION: Nucleic Acids and Methods of Use
; NUMBER OF SEQUENCES: 110
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP


```

; INFORMATION FOR SEQ ID NO: 3:
;-----
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 1251 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-02251-3

```

Query Match	2.7%	Score 155.5;	DB 5;	Length 1251;
Best Local Similarity	20.2%	Pred. No. 5.4e-05;		
Matches 149;	Conservative	68;	Mismatches 245;	
			Indels 277;	Gaps 48;

```

1  TITLE OF INVENTION:  And Tissues
2
3  NUMBER OF SEQUENCES:  3
4
5  CORRESPONDING ADDRESSES:
6
7  ADDRESSEE:  Arnold, White & Durkee
8
9  STREET:  P.O. Box 4433
10
11  CITY:  Houston
12
13  STATE:  TX
14
15  COUNTRY:  USA
16
17  ZIP:  77210
18
19  COMPUTER READABLE FORM:
20
21  MEDIUM TYPE:  Floppy disk
22
23  COMPUTER:  IBM PC compatible
24
25  OPERATING SYSTEM:  PC-DOS/MS-DOS
26
27  SOFTWARE:  Patentn Release #1.0, Version #1.30B
28
29  CURRENT APPLICATION DATA:
30
31  APPLICATION NUMBER:  US/08/199,780
32
33  FILING DATE:  18-FEB-1994
34
35  CLASSIFICATION:  514
36
37  ATTORNEY/AGENT INFORMATION:
38
39  NAME:  Parker, David
40
41  REGISTRATION NUMBER:  32,165
42
43  REFERENCE/DOCKET NUMBER:  UMIC:002
44
45  TELECOMMUNICATION INFORMATION:
46
47  TELEPHONE:  (512) 320-7200
48
49  TELEFAX:  (512) 474-7577
50
51  INFORMATION FOR SEQ ID NO:  3:
52
53  SEQUENCE CHARACTERISTICS:
54
55  LENGTH:  1252 amino acids
56
57  TYPE:  amino acid
58
59  TOPOLOGY:  linear
60
61  MOLECULE TYPE:  protein
62
63  US-08-199-780-3

```

Query Match	2.7%	Score 155.5;	DB 1;	Length 1252;
Best Local Similarity	20.2%;	Pred. No. 5.5e-05;		
Matches 149;	Conservative	68;	Mismatches 245;	
			Indels 277;	Gaps 48;

```
QY 743 GSDDTNLVGAFCVQSTIIPSEKGFRAALSSQSIILADTFIGTVVETTLKNINIKEDMF 802
Db 1037 -----LPCACT-PPAEYSPAQA-----QCLI-----PERM 1060
QY 803 PVPSTQIPDVHFFYKSTATTSCING-----RSTAVKMRNP-TKSGAGVI 847
Db 1061 STPQRDVKA---GASERTACVWGPWAGPALTFDDCCCRQRLGTQCRPCPPRGTG-- 1114
QY 848 SVPSKCPACTCGCTFYFLWESAEACPLC-----TEHDFHELEGACKRGFOETLYVWNE 901
Db 1115 ---SQCTPSQSESNF---WDTIS---PLLGLKSPRDESDSEDSDECR----- 1153
QY 902 PKWCIKIGISLPEKKLATCE 920
Db 1154 ---CVSGPCVPRPGGAVCE 1169

RESULT 9
US-08-316-650-3
; Sequence 3, Application US/08316650
; Patent No. 5942496
; GENERAL INFORMATION:
; APPLICANT: Bonadio, Jeffrey
; APPLICANT: Roessler, Blake J.
; APPLICANT: Goldstein, Steven A.
; APPLICANT: Lin, Wushan
; TITLE OF INVENTION: METHODS AND COMPOSITIONS
; FOR STIMULATING BONE CELLS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/316,650
; FILING DATE: 30-SEP-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/199,780
; FILING DATE: 30-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UMIC:008
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1252 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-316-650-3

Query Match 2.7%; Score 155.5; DB 2; Length 1252;
Best Local Similarity 20.2%; Pred. No. 5.5e-05;
Matches 149; Conservative 68; Mismatches 245; Indels 277; Gaps 48;
QY 289 ECFPCPKP--GTFSNKPGSFNCQVCP--NTYSEKGAKECIRCKDDSGSSECTERPPC 344
Db 601 EAPCGPGKGMTCMNTGCSYNCH--CNRGYRLHVGAGRSVC-----DLNECAKPHLC 650
QY 345 TTKDY---FQIHPCDEEGTQIMY-----KWIEPKICREDLTDAILRP---PSGEKKDC 393
```

```
Db 651 GDGFCINFPGHYKCN-----CYPGYRLKASRPPIK-EDI-DECRDSTPCDGKCNK 701
QY 394 PCNPGFYNGSSSCHPCPPGTFSOGTKECR---PCPAGTEPALGFYKWNVNLPGNMT 450
Db 702 P-----GSKFCTACQPGYRSQGGGACRDVNECSEGTCPSPG---WCBKLPGSYHC 748
QY 451 SC-----FNVGNSKC---DGMNGWEVAGDHQSGAGGSDNDYLIILMLHTPGFKPPTSMTG 502
Db 749 -TCAQGIKRTGRLSCIDVDCEAGKVCQDGICTNTPGSPQCOCCLSGYHL----- 797
QY 503 ATGSELGRITFV---FETPLC--SADCVLYFMVDINKSTNNVESW-----GGTK 546
Db 798 --SRDRSRCEIDECDFPAACIGDCI-----NTNGSYRCLCPLGHLRLVGGRK 843
QY 547 -----EQAVTHIIFKNATFTT---WAFORTNOGDNRR-FIN- 581
Db 844 CKKIDECSDQDGLCLPHACENLQGSYVCVDEGFTLTQDQHGCEEVEPHHKKECYLNP 903
QY 582 -DMVKIYSITATNAVDGVASSCRACALGSEOSGSSC--VPC-----PPGHYTE 626
Db 904 DDTVFCDSVLATNVQ---QEC-CCSLGAGW-GDHCEIYPCVPYSSABFIHSLVDPGKRLH 958
QY 627 KETNCKECKPPTYLSIHO--VYKREACI--PCGPGSKNNDHSVCYSCFFYHEKENOI 582
Db 959 SGQQHCELCIP-AHRDIDECILFGAEICKEGKC---VNSQPGYECYCKOGFY----- 1006
QY 683 LHYDFSNLSSVGLMNGPSTSKGTYFFHFNISLCGHGKGMALCTNITDFTVKEIVA 742
Db 1007 --YDGNLLECV-----DVDECLDES-----CRNGVCENTWR--- 1036
QY 743 GSDDTNLVGAFCVQSTIIPSEKGFRAALSSQSIILADTFIGTVVETTLKNINIKEDMF 802
Db 1037 -----LPCACT-PPAEYSPAQA-----QCLI-----PERM 1060
QY 803 PVPSTQIPDVHFFYKSTATTSCING-----RSTAVKMRNP-TKSGAGVI 847
Db 1061 STPQRDVKA---GASERTACVWGPWAGPALTFDDCCCRQRLGTQCRPCPPRGTG-- 1114
QY 848 SVPSKCPACTCGCTFYFLWESAEACPLC-----TEHDFHELEGACKRGFOETLYVWNE 901
Db 1115 ---SQCTPSQSESNF---WDTIS---PLLGLKSPRDESDSEDSDECR----- 1153
QY 902 PKWCIKIGISLPEKKLATCE 920
Db 1154 ---CVSGPCVPRPGGAVCE 1169

RESULT 10
US-08-644-271-30
; Sequence 30, Application US/08644271
; Patent No. 5814478
; GENERAL INFORMATION:
; APPLICANT: Valenzuela, et al.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS
; AND LIGANDS
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Regeneron Pharmaceuticals, Inc.
; STREET: 777 Old Saw Mill Road
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; FILING DATE: US/08/644,271
; CLASSIFICATION: 435
```

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: USSN 60/008,657
 FILING DATE: 15-DEC-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Cobert, Robert J
 REGISTRATION NUMBER: 36,108
 REFERENCE/DOCKET NUMBER: REG 195A
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 914-345-7400
 TELEFAX: 914-345-7721
 TELEX:
 INFORMATION FOR SEQ ID NO: 30:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1940 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 FEATURE:
 NAME/KEY: Rat Agrin
 LOCATION: 1...1940
 OTHER INFORMATION:
 US-08-644-271-30

Query Match 2.7%; Score 153; DB 2; Length 1940;
 Best Local Similarity 17.9%; Pred. No. 0.00019;
 Matches 174; Conservative 86; Mismatches 302; Indels 408; Gaps 49;

```

QY 219 IONDOCEMDTTDKWKLTDNGENGSH-----SYMLKSGTNIL-----Y 258
Db 118 LQAOQCNQ-----QRRILLRQPGCGSRDPCANVTCSFGSTCVPSADGQTASCLCPTTCF 172
QY 259 WRTTGILMGSKAVKPVLVKNITIEGVATYSEC-----PPCKPGTGSN 300
Db 173 GAPDGTVCGS-----DGVDPSECOLLSHACASOEHIFFKFKENGPCDPCQGSN 219
QY 301 KPGSFNCVCPNRTYSEK-----GAKECIRCKDSDS-----QFS 333
Db 220 SDLNHICRVNPRTRHPENMLLRPENCAPGHTPICGDDGVTYENDCVMSRIGATRGLLQKV 279
QY 334 GSSECTERPPCTTKDYF-----QIHTPCD-----EBGKTQIMYKWIETP 371
Db 280 RSGQOTRDQCQFNSVCLSRGRPHCSRDVTCDSYRVPVCAQDGHYNNDCWRQ 339
QY 372 KICREDLTDALRLPPSGEKKDCPCPNPCFYNNNGSSSCH--PCPPG---TFSDGTKEC--- 423
Db 340 AECR-----QORAIIPKHQGPCDQTPSPCHGVQCAFCAVCTVRNGKAECEQ 386
QY 424 RPCPAGTEPALGFYKWNVLPNNKTSCTFNVGNSKCDGMNGWEVAGDHIQSGAGGSND 483
Db 387 RVCSGIYDPVCGSD-----GVTVGSVCELESMAAC-----TLGREIQVARRG--- 427
QY 484 YLILNLHIPGFKPPTSMTGAT--GS-----ELGRITTFVETLCSADCVLYFWVDINRKTN 537
Db 428 -----PCDPCGQCFRGLSCEVETGR-----CVCPSCEV-----ESAOP 460
QY 538 VVESGGTKEKOAYTHIIFKNATFTTWFORTNOGODNRFRINDMVKIYSITATNAVVDG 597
Db 461 VCGSDGHTYASECELHV-----HACTHOISLYASA-----G 492
QY 598 VASSC--RACALGSQSGSSVCPGPGHYIEKTNQCKECPDP-----TYLSI----- 643
Db 493 HCOTGCKVCTFGVACSAGQCV-CP-----RCEHPPPPVPGVCGSDGVTYLSACELR 541
QY 644 -----HQVYKGEA-----CIP--CGPGSKNNODSHSVCYSD-C-----PFYHEKENOILHYD 686
Db 542 EAACQOQVIEAHAGCEPAECSSGGSGEDDECEBELCRQGGIWDSEDPGCVCD 601
QY 687 FSNLSSVGLMNGPSFTSKGTYKTHFFNISLCGHGKMAALCTNNITDFTVKEIVAGSD 746
Db 602 F-----SCOSVPRSP-----VCGSDGVTYG-----TECDLKKARCESQ 635
QY 747 YTNLVGAFVCS--TIIPSEKGFRAALSSQSIIILADTFIGVTVETTLKINIKEDMFVPV 805
  
```

Db 636 ELYVAAQACRGPTLAP-----LLPVA 657

```

QY 806 TSQIPDVHFFYKSTATTSCINGRSTAVKMRNPTKSGAGVISVPSKCP-----AGTC 858
Db 658 -----FPHCAQTPYGCCODNNTAAQ-----GVGLAGCPSTCHCNPHGYSYSGTC 700
QY 859 DGCT-----FY-----FLWESAECPLCT-----EHDFFEIEGA 887
Db 701 DPATGQCSRPGVGGRLCRDRCPEGFNFRGIVTDHSGCTPCSDPRGAVRDDCEQMTGL 760
QY 888 CKRGFOETLYVMNEPKWCLIGISLPEKKLATCETVDFWMLKVGAGVGAFTAVLLVALTVCVF 947
Db 761 CS-----CRPGVAGP--KCGQCPDQGVGLHGLCEADPMTPTVCVEIHCEF 803
QY 948 WKKNOKLEYKSKLYMTTNSKEC-----ELPAADSCAIMEGEDNEEEVYVSNKQSLILGKL 1002
Db 804 G-----ASCVEKAGFAQCICPTLTCTPEANSTKVC-GSDG---VTYGNE---COL 845
QY 1003 KSLATKEKED 1012
Db 846 KAIACRQRILD 855
  
```

RESULT 11
 US-09-077-955-34
 Sequence 34, Application US/09077955A
 Patent No. 6413740
 GENERAL INFORMATION:
 APPLICANT: Valenzuela et al., David M.
 TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
 FILE REFERENCE: REG195-B-PCT-US
 CURRENT APPLICATION NUMBER: US/09/077,955A
 EARLIER FILING DATE: 1998-09-10
 EARLIER APPLICATION NUMBER: PCT/US96/20696
 EARLIER FILING DATE: 1996-12-13
 EARLIER APPLICATION NUMBER: 08/644,271
 EARLIER FILING DATE: 1996-05-10
 EARLIER APPLICATION NUMBER: 60/008,657
 EARLIER FILING DATE: 1995-12-15
 NUMBER OF SEQ ID NOS: 36
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 34
 LENGTH: 1940
 TYPE: PRT
 ORGANISM: Rattus sp.
 US-09-077-955-34

Query Match 2.7%; Score 153; DB 4; Length 1940;
 Best Local Similarity 17.9%; Pred. No. 0.00019;
 Matches 174; Conservative 86; Mismatches 302; Indels 408; Gaps 49;

```

QY 219 IONDOCEMDTTDKWKLTDNGENGSH-----SYMLKSGTNIL-----Y 258
Db 118 LQAOQCNQ-----QRRILLRQPGCGSRDPCANVTCSFGSTCVPSADGQTASCLCPTTCF 172
QY 259 WRTTGILMGSKAVKPVLVKNITIEGVATYSEC-----PPCKPGTGSN 300
Db 173 GAPDGTVCGS-----DGVDPSECOLLSHACASOEHIFFKFKENGPCDPCQGSN 219
QY 301 KPGSFNCVCPNRTYSEK-----GAKECIRCKDSDS-----QFS 333
Db 220 SDLNHICRVNPRTRHPENMLLRPENCAPGHTPICGDDGVTYENDCVMSRIGATRGLLQKV 279
QY 334 GSSECTERPPCTTKDYF-----QIHTPCD-----EBGKTQIMYKWIETP 371
Db 280 RSGQOTRDQCQFNSVCLSRGRPHCSRDVTCDSYRVPVCAQDGHYNNDCWRQ 339
QY 372 KICREDLTDALRLPPSGEKKDCPCPNPCFYNNNGSSSCH--PCPPG---TFSDGTKEC--- 423
Db 340 AECR-----QORAIIPKHQGPCDQTPSPCHGVQCAFCAVCTVRNGKAECEQ 386
QY 424 RPCPAGTEPALGFYKWNVLPNNKTSCTFNVGNSKCDGMNGWEVAGDHIQSGAGGSND 483
  
```

Db 367 RVCSGIYDPCGSD-----GVYGSVCBLESMAC-----TLGREIQVARRG----- 427
Qy 484 YLILNLHPGKPTMTGAT--GS-----ELGRITFTFETLCSADCVLVFMVDINRSTN 537
Db 428 -----PCDPCGQCRFGSLCEVETGR-----CVCPCSECV-----ESAQP 460
Qy 538 VVESGGTKEQAVTHILFKNATFTTFAFQRTNQGDNRFRINDMVKIYSITATNAVVG 597
Db 461 VCGSDGHTYASECELHV-----HACTHOISLYASA-----G 492
Qy 598 VASSC--RACALGSQSGSSCVPCPGHYIEKETNQKCEPPD-----TYLSI----- 643
Db 493 HCQTGCKVCTFGVCSAGOCV-CP-----RCEHPPPGPVCGSDGVTYLSACELR 541
Qy 644 -----HQVYGEA-----CIP--CGPGSKNNODHVCYS-D-C-----FFYHEKENQILHYD 686
Db 542 EAACQOQVIEEHAHAGPCEPAECGSGSGGDEDECEQELCRQRGIWDESDGPCVCD 601
Qy 687 FSNLSSVGLMNGPSTSKGTKYHFFNLSLCGHEGKKMALCTNNITDFTYKEIVAGSD 746
Db 602 F-----SCQSVPRSP-----VCGSDGVTYG-----TECDLKKARCESQ 635
Qy 747 YTNLVGAFVQGS--TTPESKGFRAALSSQSIIILADTFIGVTVTTLKNINIKEDMFPVP 805
Db 636 ELYVAAQACRGPTLAP-----VCGSDGVTYG-----LLPVA 657
Qy 806 TSQIPDVHFFKSSATTSCINGRSTAVKMRNCPKSGAGVISVPSKCP-----AGTC 858
Db 658 -----FPHCAOTPYGCCODNFTAAQ-----GVGLAGCPTCHNPHGYSYSGTC 700
Qy 859 DGCT-----FY-----FLWESAACPLCT-----EHDFFEIEGA 887
Db 701 DPATGQCSRPGVGLRCDCEPGFNFNRIIVTDGHSCTPCSCDPRGAVRDDCEQMTGL 760
Qy 888 CKRGQETLYVWNEPKWICKISLEPKKLATCETVDFWLKVGAGVGAFTAVLLVLTCTYF 947
Db 761 CS-----CRPGVAGP--KCGQCPDQVVLGHLGCEADPMTPTVTCVEIHCF 803
Qy 948 WKKNQKLEYKSLVMTNNSKEC-----ELPAADSCAIMEGEDNEEVVYSNKQSLGLKL 1002
Db 804 G-----ASCVEKAGFACICPTLTCPEANSTKVC-GSDG-----VITYGNE-----COL 845
Qy 1003 KSLATKEKED 1012
Db 846 KAIACRQRLD 855

RESULT 12
US-08-718-388-7
Sequence 7, Application US/08718388
Patent No. 6271362
GENERAL INFORMATION:
APPLICANT: MORIKAWA, MINORU
APPLICANT: HARADA, NAOKI
TITLE OF INVENTION: GENE ENCODING Igg Fc REGION-BINDING
TITLE OF INVENTION: PROTEIN
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
STREET: PO BOX 747
CITY: FALLS CHURCH
STATE: VA
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/718,388
FILING DATE:

CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: MURPHY JR, GERALD M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 0230-111
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2594 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-718-388-7
Query Match 2.7%; Score 151; DB 4; Length 2594;
Best Local Similarity 17.5%; Pred. No. 0.00047;
Matches 202; Conservative 103; Mismatches 346; Indels 502; Gaps 62;
Qy 18 AEAPRRGRSPWPSPAWICCWALAGCAAWAGDLPPSS-----SRPLPPCQEKDYHFEYT 71
Db 765 AAPSNCSGRCEVCVC---LPGFVASGGACVPASSCGCTFQGLQLAPGQEV----- 814
Qy 72 ECDSSGSRWRVAIPNSAVDCSGLDDPV--ROKECTFCASGEYLEMKNOV----- 119
Db 815 -----WADLCORRCTCGATHQVTCRDQ---SCPAGERCSVQNGLLGCYPDRFGT 863
Qy 120 CSKCGEGTSLGSGIKFDEWDELPAFESNIATFMDTVVGVSDSRPDGCGNNSWIPRGNYI 179
Db 864 CGSGDHPHYVDFGRFD-----FMGTCTYL--LVG-----SCQNAALPAFRLV 906
Qy 180 ESN--RDCTVTSIYAVHLKSGY---VFVEYQ---YVDNNI--FFEFTIQDQCEMDYT 230
Db 907 VENEHRGSGTYSYTRAVRVEARGVAVRREYVPGQVLDVDDVQLPFOAAGQVQVPRQG 966
Qy 231 TDKWKV-----LTDNGEWGSH--SVMLKSGTNILYRTTGLMGSKAVKPVL-----VKNT 280
Db 967 RDVAVRTDGLTVDNARVTAKYVSSYAEALCGLCNFNGDPADDLALRGGAANAL 1026
Qy 281 EGVAVTSECPCKPGTFSNKPSPENCQVCP-----CPKLDLSVLAQQLSKNECGLADPKGPFREC 314
Db 1027 AFGNSWQEE---TRPGCGATEPGD-----CPKLDLSVLAQQLSKNECGLADPKGPFREC 1078
Qy 315 YSE---KGA-KECI--RCKDDSQFSSECTERPPCTTKDYFIHTPDCEKTKTQIMKW 368
Db 1079 HSKLDPOGAVRDCVYDRCLLPQ--SG-----PLCDALATYA--AACQAAGAT--VHPW 1126
Qy 369 IEPKICREDLTDAILRPPSGEKKDCP-----PCNPGFYNNG 404
Db 1127 RSEELC-----PLSCPPIISHYEACSYGCLSCGDLVPVGGCGSECHGECVCEGFAISG 1180
Qy 405 SSSCHP-----C-----PPG--TF-----SDGTEKCRPCPAGTEPALGFEY 438
Db 1181 -ESCLPLASCGCVHOGTYHPPGQTFYPGPCDSLCHCQSGGLVSCSESSCGPIEA----- 1234
Qy 439 KWNVLPGNMKTSCTNVGNSKCDGMNGVEVAGDHTQSGAGSDNDYLIILNLHIPGKPT 498
Db 1235 ----CQPSGSGLCVAVGSSTC-----QASGD-----PHY--T 1261
Qy 499 SMTGATGSELGRITVFETLCSADCVLYPMVDINRKTNNVSVESWGTKEKQAVTHIL-PE 557
Db 1262 TFDGRRFDFMGTCTVYLAQTCGTRGLHIFRAVLQE---NV--AWG--NGRVSVTRVITVQ 1314
Qy 558 NATFTF-----TW-----AFORTNQGD-----NRRFINDMVKIY 587
Db 1315 VANFTLRLRQKQKVTVNGVDMKLPVVLANGOIRASQHSDDVVIETDEGLRAYDLV--Y 1372
Qy 588 SITATNAVGVASSCRACA-----LGSQSG-----SSCV--- 617
Db 1373 YVRVTVPGNYVQCMGLCGNYNGDPKDDFKPKNGSOAGNANEFNGSWEEVVPDSCPPT 1432
Qy 618 PCPPGHIYEKTNQCKECPDITYLSIHVYVGEACIPCGPGSKNNQDHSVCY----- 669

Db 1433 PCPPG---SDDCIPSHKCPPE-----LEKRYKEEF--CGLLSPTGPLSSCHKLVDPQGP 1483
QY 670 -SDCFVHEKENQILHYDFSNLSSVGLMNGPSTSGTKYFHFFNLSLCGHEGKMAIC 728
Db 1484 LKDCIF-----DLCGGNLSILC 1502
QY 729 TNNITDFTVKEIVAGSDDYTNLVGAFV--COST---IIPSEKGFRAALSSOSIILADTF 783
Db 1503 SN-----IHAVVSACQAAGHVEPWRTETF-----1527
QY 784 IGVTVETTLKNINIKEDMPVPVTSQIPDVHFFYKSTATTSCINGRSTAVKMRCPNPKSG 843
Db 1528 -----CPMECPNPNHYELCADTCSILG-----1549
QY 844 AGVISVPSKCPAGTCDG--CTFYFLWESAEACPL-----875
Db 1550 -SALSAPPQCGCAGCQCDSGFLYNGQACVPICQCGCYHNGVYVEPQTVLIDNCROQ 1608
QY 876 -----CTEHD-----FHEI-----EGACKRG 891
Db 1609 CTCHAGKGMVCOEHSCKPGQVCPGSGILSCVTKDKPCHGVTCRPQETCKEQGGVCLPN 1668
QY 892 FQETLVWNEPKW 904
Db 1669 YEATCWLGDPHY 1681

RESULT 13

US-08-718-388-9
Sequence 9, Application US/08/18388
Patent No. 6271362
GENERAL INFORMATION:
APPLICANT: MORIKAWA, MINORU
APPLICANT: HARADA, NAOKI
TITLE OF INVENTION: GENE ENCODING I9G FC REGION-BINDING
TITLE OF INVENTION: PROTEIN
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
STREET: PO BOX 747
CITY: FALLS CHURCH
STATE: VA USA
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/718,388
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: MURPHY JR, GERALD M
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 0230-111
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 205-8000
TELEFAX: (703) 205-8050
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 5405 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-718-388-9

Query Match 2.7%; Score 151; DB 4; Length 5405;
Best Local Similarity 17.5%; Pred. No. 0.0016;
Matches 202; Conservative 103; Mismatches 346; Indels 502; Gaps 62;

QY 18 AEAPRRGRSPWPSPATWCWALAGCOAAGADLPSS-----SRPLPPCOEKDYHFEYT 71
Db 765 AAAPNSCNRPCVGEVCV---LPGFVASGACVPASCGCTFOGLQLAPQGEV-----814
QY 72 ECDSSGRMRVAIPNSAVDCSGLPDPV--RGKECTSCASGEYLEMKNOV-----119
Db 815 -----WADELQCRCTCNGATHQVTCRDQO---SCPAGERCSVQNGLLGCYDPDRFGT 863
QY 120 CSKCGEGTYSLSGIKFDWDELPAFESNIATFMDTVVGPDSRSDPGCCNNSSWIPRGNYI 179
Db 864 CQSGSDPHVYSFEDGRFD-----FMGTCTYLV--LVG-----SCGQNAALPAFRVL 906
QY 180 ESN--RDDCTVSLIYAVHLKSGY---VFEEYO---YVDNNI--FEEFFIQNDCCQEMDIT 230
Db 907 VENEHRCSTVSTRVAREARGVKVAVREYPGQVLVDVLIQYLPFOADGOVQVFRQG 966
QY 231 TDKWKV-----LTDNGRWGSH-SVMLKSGTNILYWRITGILMGSKAVKPVL-----VKNIT 280
Db 967 RDAVVRTDFGLTVYDWNARVTAKVPSYAEALCGLCGNFNGDPADDLALRGGOAANAL 1026
QY 281 IEGVAYTSECFCKPGTFSNKPQSFNCQVCP-----RNT 314
Db 1027 AFGNSWQEE---TRPGCGATEPGD-----CPKLDLSVAQQLQSKNECGILADPKGPFREC 1078
QY 315 YSE---KGA-KECI--RCKDDSOFGSSSECTERPPCTTKDYFOIHTPCDEEGKTOIMYKW 368
Db 1079 HSKLDPOGAVRCVYDRCLLPQO--SG-----PLCBALATYA--AACQAAGAT--VHPW 1126
QY 369 IEPKICREDLTDAILRPPSGEKKDCP-----PCNPGFYNG 404
Db 1127 RSEELC-----PLSCPPIHSHVEACSYGCPPLSCGDLVPVPGCGSECHGCVCDGEFALSG 1180
QY 405 SSSCHP-----C-----PPG-TF-----SDGTECRPCPAGTEPALGFY 438
Db 1181 -ESCLPLASCGCVHOGTYHPPGQTFYPPGCDLSLCHQCEGLVSCSESSCGPHEA-----1234
QY 439 KWNVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIOGAGGSDNDYLIILNHLHPGKPPPT 498
Db 1235 ----CQPSGSLGCVAVGSGTC-----QASGD-----PHY---T 1261
QY 499 SMTGATGSELGRITFVFETLCSADCVLYFMVDINRKSTNVVSWGCTKEQAYTHII-FK 557
Db 1262 TFDGRRDFMGTCVYVLAOTCGTRPGLHRAVLQE---NV--AWG--NGRVSVTRVITVQ 1314
QY 558 NATETF---TW-----AFORTNQOD-----NRRFINDMVKIY 587
Db 1315 VANFTLRLEQROWKVTNVGDMKLPVVLANGOIRASQHGSDVVIETDFGLRVAIDLVL--Y 1372
QY 588 SITATNAVDGVASSCRACA-----LGSEQSG-----SSCV-----617
Db 1373 YVRVTVPGNYYOQMCGLCGNYNGDKDDFKPNGSQAGNANEFGNSWEVVPDPSCLPPT 1432
QY 618 PCPPGHIYEKETHNOCKECPDPTYSLIHQVYGREACIPCGSKNNODHVSVCY-----669
Db 1433 PCPPG---SEDCIPSHKCPPE---LEKRYKEEF--CGLLSPTGPLSSCHKLVDPQGP 1483
QY 670 -SDCFVHEKENQILHYDFSNLSSVGLMNGPSTSGTKYFHFFNLSLCGHEGKMAIC 728
Db 1484 LKDCIF-----DLCGGNLSILC 1502
QY 729 TNNITDFTVKEIVAGSDDYTNLVGAFV--COST---IIPSEKGFRAALSSOSIILADTF 783
Db 1503 SN-----IHAVVSACQAAGHVEPWRTETF-----1527
QY 784 IGVTVETTLKNINIKEDMPVPVTSQIPDVHFFYKSTATTSCINGRSTAVKMRCPNPKSG 843
Db 1528 -----CPMECPNPNHYELCADTCSILG-----1549
QY 844 AGVISVPSKCPAGTCDG--CTFYFLWESAEACPL-----875
Db 1550 -SALSAPPQCGCAGCQCDSGFLYNGQACVPICQCGCYHNGVYVEPQTVLIDNCROQ 1608
QY 876 -----CTEHD-----FHEI-----EGACKRG 891

Db 1609 CTCHAGKGMVCOEHSCKPQVQCPQSGGILSCVTKDPCHGVTCRPOETKEQGQGVCLPN 1668
Qy 892 FOETLYVWNEPKW 904
Db 1669 YEATCWLWMDPHY 1681

RESULT 14
US-08-185-432-18
; Sequence 18, Application US/08185432
; Patent No. 5750652
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon
; APPLICANT: Busseau, Isabelle
; APPLICANT: Diederich, Robert J.
; APPLICANT: Xu, Tian
; APPLICANT: Matsuno, Kenji
; TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/185,432
; FILING DATE: 21-JAN-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2523 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-185-432-18

Query Match 2.6%; Score 149.5; DB 1; Length 2523;
Best Local Similarity 19.0%; Pred. No. 0.00062;
Matches 168; Conservative 83; Mismatches 284; Indels 349; Gaps 50;

Qy 278 NITIEGVAVTSEC--EPCK-PGTFSNKPQSGFNCQVCPNRY-----SEKAKECI-- 324
Db 675 NINI-----NECDSPCRNGGTCKQDQINGFTC-VCPDGYHDMCLSEVNECSNPCIHG 727
Qy 325 -----RCKDDSOFGSGS-----SECTERP-----PCTTKDYFO----- 351
Db 728 ACHDGVNGYKDCDEAGWSGSGNDINNECESNPMNGGTCKDMTGAYICTKAGFSGPNC 787
Qy 352 -----IHTPCDEEKTQIMYKWIPEKTKRED-----LTDAILRPPSGE 389
Db 788 QTNINECSNPNLHG-----TCIDDDVAGYKCNCLPPTYGAICEAVLAPCAGS 835
Qy 390 KKDCCPPCNPGFYNGSS-----SCHPCPPG-----TESDGTKEC--RPPCPAGTEPALGFEY 438
Db 836 -----PCKNGGRKESDEFTFSCC-CPFGWQQTCEIDMNECVNRCRNG----- 880

Qy 439 KWNVLPNGMKTSCFNV-GNSKCDGMNGH-----EVAGDHTQSGA---GGSDNDYLIILNL 489
Db 881 -----ATCQNTNGSYKCNCKPGVYTRNCNEMDIDDCQPNPCHNGGSGSDG--INM 927
Qy 490 HI-----PGFKPPTSMTGATGSELGRITVFETLCSADCVLYFMVDINRSTNNVSWGCT 545
Db 928 FFCNCPAGFRGPK-----CBE-----DINECASNPCKNGA-- 957
Qy 546 KEKQAYTHIIFKNATFTTFAFORTNQGDNRREINDMVKIYSITATNAVDCV--ASSCR 603
Db 958 -----NCT-----DCVNSYTCQCPQFSGIHCESNTP 984
Qy 604 ACALGSEGGSSCV-----PCPP--GHYIEKETNOCKECPDPTLYSIHQVYKKEAC 652
Db 985 DCTESSCFNGGTCDIGINTFTCCQPPGFTGSGYCOHIDNECDKPLNGGTCQDSYCTYKC 1044
Qy 653 I-----PCGPGSKNNODHSVCYSDFYHEKENQLHYDFSNLSS 692
Db 1045 TCPQGYTGLNCONLVKMDSSPKNGKGKQWQNNFYRCEC-----KSGW 1088
Qy 693 VGLSLMNGPSFT-----SKTRYFHF--NISLC-----GHEGKKMALCTNNI 732
Db 1089 TGVYCDVPSVCEVAAKQGVQVDIVHLCRNSGMCVDTGNTHECRQAGYTG---SYCEGV 1145
Qy 733 TDTVKEIVAGDDYTNLVGAFCVOSTIIPSEKGFRAALSSQSILADTFILGVIVETTL 792
Db 1146 DECSNPNCNGA--TCTDYLGGYSCCV-----AGVHGVNCSSEINECLSHPCQNGGTCT 1198
Qy 793 KNINIKEDMFVPYPTS-----QIPDVHFFYKSTATTSCINGRSTAVKMRNCTKSGAGV 846
Db 1199 DLINTYKSCPRGTQGVHCEINVDCTPEYDSFTLEPKCFNNG-----KCIDRVGGYNC 1252
Qy 847 ISVPSKCPAGTCDGCTTFYFLWESAPACPLCTEHPHE-IEGAC-KRGEQETLYVWNEPKW 904
Db 1253 I-----CPPG-----DVNECLSNPCDSRGTCQICQLVNDYRC 1292
Qy 905 -CIKGIS-----LPEKKLATCETVD-----FMLKVAGAGVGAFTAVLLVALTC 945
Db 1293 ECRQGTGRRRCESVVDGCKGMPCRNGGTCAVASNTERGICRPPGFDGATGE-VDSRTC 1351
Qy 946 YFWKNQKLEYKYKSLVMTTNSK-----ECLEPAADSCA 979
Db 1352 ----SNLRQNGGTCLISVLTSKVCSEGYTGATCQYVISPCA 1391

RESULT 15
US-08-899-232-3
; Sequence 3, Application US/08899232
; Patent No. 6436650
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon
; APPLICANT: Qi, Huilin
; TITLE OF INVENTION: ACTIVATED FORMS OF NOTCH AND METHODS BASED THEREON
; FILE REFERENCE: 7326-046
; CURRENT APPLICATION NUMBER: US/08/899,232
; CURRENT FILING DATE: 1997-07-23
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2523
; TYPE: PRT
; ORGANISM: Xenopus sp.
US-08-899-232-3

Query Match 2.6%; Score 149.5; DB 4; Length 2523;
Best Local Similarity 19.0%; Pred. No. 0.00062;
Matches 168; Conservative 83; Mismatches 284; Indels 349; Gaps 50;
Qy 278 NITIEGVAVTSEC--EPCK-PGTFSNKPQSGFNCQVCPNRY-----SEKAKECI-- 324
Db 675 NINI-----NECDSPCRNGGTCKQDQINGFTC-VCPDGYHDMCLSEVNECSNPCIHG 727
Qy 325 -----RCKDDSOFGSGS-----SECTERP-----PCTTKDYFO----- 351

Db 728 ACHDGVNGYKDCGAGWSSGNCIDINNCESNPCMGNGTCKDMTGAYICTCKAGFSGPN 787
QY 352 -----IHTPCDEBKQIMYKWIPIKICRED-----LTAIRLPPSGE 389
Db 788 QTNINESSNFCNLNHG-----TCIDDVAGYKCNMLPYTGAICEAVLAPCAGS 835
QY 390 KKDCPPCPNPGFYNGSS-----SCHPCPPG---TFSDGTKEC--RPGPAGTEPALGFEX 438
Db 836 -----PCKNGGRKESDEFETFSCE-CPPGWOGQTCEIDMNECVNRPCNG----- 880
QY 439 KWNVLPGNMKTSFNV-GNSKDCGMNGW-----EVAGDHIQSGA---GGSDNDYLIINL 489
Db 881 -----ATCQNTNGYKCNCKPGYTGRCNCEMDIDCQPNPCNHNCGSCSDG--INM 927
QY 490 HI-----PGFKPPTMTGATGSELGRITVFETLCSADCVLFWVDINRKSNNVSVESWGT 545
Db 928 FFCNCPAGFRGPK-----CEE-----DINECASNPCKNGA-- 957
QY 546 KEQAYTHIIFKNATFTFWAFORTNOGDNRRFINDMVKIYSITATNAVDGV--ASSCR 603
Db 958 -----NCT-----DCVNSYTCQCPGFSGIHCESNTP 984
QY 604 ACALSGESGSSCV-----PCPP---GHYIETKNOCKECPDPTYLSTHQVYVKEAC 652
Db 985 DCTESSCFNGGTCIDGINTFTCQCPGPGTGYSCQHDINECDKPCNLNGGTQODSYGTYKC 1044
QY 653 I-----PCGPGSKNNQDHSVCYSDCFYHEKENOILHYDFSNLSS 692
Db 1045 TCPQGYTGLNQNVLVRWCDSSPCKNKGRCWQTNNFYRCE-----KSGW 1088
QY 693 VGSLLMNGPSFT-----SKGTKYHFF--NISLC-----GHEGKKMALCTNNI 732
Db 1089 TCVYCDVFSVCEVAAKOQGVDIHLCRNSGMCVDTGNTFCRCQAGYTG---SYCEEQV 1145
QY 733 TDFTVKEIVAGSDYTNLVGAFVCOSTIIPSESKGFRAALSQSIIADTFIGVTVETTL 792
Db 1146 DECSNPQCNGA-TCTDYLGGYSCBV-----AGYHGVNCSEEEINECLSHPCQNGGTCTI 1198
QY 793 KNINIKEDMFVPPTS-----QIPDVHFFYKSSATTSCINGRSTAVKMRNCPNPKSGAGV 846
Db 1199 DLINTYKCSCPRGTOGVHCEINVDCTPFYDSFTLLEPKCFNNG-----KCIDRVGGYNC 1252
QY 847 ISVPSKPCAGTCDCGTFYFLWESAEACPLCTEHPHE-IEGAC-KRGFOETLYVWNEPKW 904
Db 1253 I-----CPPG-----FVGECGE-----DYNECLSNPCDSRGTONCQLVNDYRC 1292
QY 905 -CIKGIS-----LPEKKLATCEIVD-----FWLKVAGAGVGAFTAVLLVALTC 945
Db 1293 ECRQGTGRRCESVVDGCKGMPCRNGGTCAVASNTERGFIKCPGPDGATCE-YDSRTC 1351
QY 946 YFWKKNOLEKYKLVMTNSK-----ECELPAADSCA 979
Db 1352 -----SNLRCONGGTCSILTSSKVCSEGYTGATCOYPISPCA 1391

Search completed: May 12, 2003, 13:23:54
Job time : 35.1915 secs


```
Qy 61 COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120
Db 61 COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120
Qy 121 SKCEGYTSLGSGIKFDEWDELPAFNSIATFMDTVVGPDSRDPDGCNNSWIPRGNYIE 180
Db 121 SKCEGYTSLGSGIKFDEWDELPAFNSIATFMDTVVGPDSRDPDGCNNSWIPRGNYIE 180
Qy 181 SNRDDCTVSLIYAVHLKSGYVFEYQYVDNNIFEFFIQNDQCEMDTTTDDKWVKLTDN 240
Db 181 SNRDDCTVSLIYAVHLKSGYVFEYQYVDNNIFEFFIQNDQCEMDTTTDDKWVKLTDN 240
Qy 241 GEMGSHVLMKSGTNILYWRITGILMGSKAVKPVLVKNITIEGVAYTSCFPCKPCTFSN 300
Db 241 GEMGSHVLMKSGTNILYWRITGILMGSKAVKPVLVKNITIEGVAYTSCFPCKPCTFSN 300
Qy 301 KPGSFNCQVCPRNTYSEKGAKECIRCKDDSOFSGSSSECTERPPCTTKDYFOIHTPCDEEG 360
Db 301 KPGSFNCQVCPRNTYSEKGAKECIRCKDDSOFSGSSSECTERPPCTTKDYFOIHTPCDEEG 360
Qy 361 KTOIMYKWIIEPKICREDLTDAILRPPSGEKKDCPPCNPFGYNNGSSSCHPCPPGTFSDGT 420
Db 361 KTOIMYKWIIEPKICREDLTDAILRPPSGEKKDCPPCNPFGYNNGSSSCHPCPPGTFSDGT 420
Qy 421 KECRCPAGTEPALGFYKKNWVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIOGAGGS 480
Db 421 KECRCPAGTEPALGFYKKNWVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIOGAGGS 480
Qy 481 DNDYLILNLHLPGRKPPPTSMGTGATGSELGRITFVFETLCSADCVLVFMVDINRKSTNVVE 540
Db 481 DNDYLILNLHLPGRKPPPTSMGTGATGSELGRITFVFETLCSADCVLVFMVDINRKSTNVVE 540
Qy 541 SWGGTKEKOAYTHIIFKNATFTTFAFORTNOGQDNRFFINDMKIYISITATNAVGVAS 600
Db 541 SWGGTKEKOAYTHIIFKNATFTTFAFORTNOGQDNRFFINDMKIYISITATNAVGVAS 600
Qy 601 SCRACALGSESGSCVPCPGHYIEKETNOCKECPDPTIYLSIHQVYKKEACICPGPGSK 660
Db 601 SCRACALGSESGSCVPCPGHYIEKETNOCKECPDPTIYLSIHQVYKKEACICPGPGSK 660
Qy 661 NNQDHSVCYSDCFYHEKENQILHYDFNSLSVGLMNGPSPFTSKGTKYFHFNFNLSLGH 720
Db 661 NNQDHSVCYSDCFYHEKENQILHYDFNSLSVGLMNGPSPFTSKGTKYFHFNFNLSLGH 720
Qy 721 EGKKNALCTNNITDPTVKEIVAGSDDYTNLVCAFVCOQSTIIPSESKGFRALSSQSIILA 780
Db 721 EGKKNALCTNNITDPTVKEIVAGSDDYTNLVCAFVCOQSTIIPSESKGFRALSSQSIILA 780
Qy 781 DTFIGTVVETTLKINIKEDMFPVPTSOIPDVHFEYKSTATTSCINGRSTAVKMRNCNPT 840
Db 781 DTFIGTVVETTLKINIKEDMFPVPTSOIPDVHFEYKSTATTSCINGRSTAVKMRNCNPT 840
Qy 841 KSGAGVISVPKCPAGTCGCTFYFLWESAECPLCTEHDHFEIEGACKRGFOETLYVWN 900
Db 841 KSGAGVISVPKCPAGTCGCTFYFLWESAECPLCTEHDHFEIEGACKRGFOETLYVWN 900
Qy 901 EPKWCIGISLPEKKLATCETVDFWLKVGAGVGAFTAVLLVALTCYFWKKNOKLEYKYSK 960
Db 901 EPKWCIGISLPEKKLATCETVDFWLKVGAGVGAFTAVLLVALTCYFWKKNOKLEYKYSK 960
Qy 961 LVMTTNSKECELPAAADSCAIMEGEDNEBEVYVSNKQSLGKLKSLATKEKEDHFESVOLK 1020
Db 961 LVMTTNSKECELPAAADSCAIMEGEDNEBEVYVSNKQSLGKLKSLATKEKEDHFESVOLK 1020
Qy 1021 TSSRSPNI 1027
Db 1021 TSSRSPNI 1027
```

RESULT 2
US-10-140-164-2
; Sequence 2, Application US/10140164
; Publication No. US20030072736A1

```
GENERAL INFORMATION:  
; APPLICANT: Baker et al;  
; TITLE OF INVENTION: Human Tumor Necrosis Factor Receptor TR16  
; FILE REFERENCE: PF514C1  
; CURRENT APPLICATION NUMBER: US/10/140.164  
; CURRENT FILING DATE: 2002-05-08  
; PRIOR APPLICATION NUMBER: 09/637,856  
; PRIOR FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/148,348  
; PRIOR FILING DATE: 1999-08-12  
; PRIOR APPLICATION NUMBER: 60/148,683  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: 60/148,870  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: 60/148,758  
; PRIOR FILING DATE: 1999-08-16  
; PRIOR APPLICATION NUMBER: 60/149,181  
; PRIOR FILING DATE: 1999-08-17  
; PRIOR APPLICATION NUMBER: 60/149,453  
; PRIOR FILING DATE: 1999-08-18  
; PRIOR APPLICATION NUMBER: 60/149,498  
; PRIOR FILING DATE: 1999-08-19  
; NUMBER OF SEQ ID NOS: 76  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 963  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-140-164-2
```

Query Match 93.4%; Score 5307; DB 9; Length 963;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 953; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 MLFRARGVVRGWRGPAEAPRRGRSPWPSPAWICCWALAGCQAAGWAGDLPPSSSRPLPP 60
Db 1 MLFRARGVVRGWRGPAEAPRRGRSPWPSPAWICCWALAGCQAAGWAGDLPPSSSRPLPP 60
Qy 61 COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120
Db 61 COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKECTFSCASGEYLEMKNQVC 120
Qy 121 SKCGGTVSLGSGIKFDEWDELPAFNSIATFMDTVVGPDSRDPDGCNNSWIPRGNYIE 180
Db 121 SKCGGTVSLGSGIKFDEWDELPAFNSIATFMDTVVGPDSRDPDGCNNSWIPRGNYIE 180
Qy 181 SNRDDCTVSLIYAVHLKSGYVFEYQYVDNNIFEFFIQNDQCEMDTTTDDKWVKLTDN 240
Db 181 SNRDDCTVSLIYAVHLKSGYVFEYQYVDNNIFEFFIQNDQCEMDTTTDDKWVKLTDN 240
Qy 241 GEMGSHVLMKSGTNILYWRITGILMGSKAVKPVLVKNITIEGVAYTSCFPCKPCTFSN 300
Db 241 GEMGSHVLMKSGTNILYWRITGILMGSKAVKPVLVKNITIEGVAYTSCFPCKPCTFSN 300
Qy 301 KPGSFNCQVCPRNTYSEKGAKECIRCKDDSOFSGSSSECTERPPCTTKDYFOIHTPCDEEG 360
Db 301 KPGSFNCQVCPRNTYSEKGAKECIRCKDDSOFSGSSSECTERPPCTTKDYFOIHTPCDEEG 360
Qy 361 KTOIMYKWIIEPKICREDLTDAILRPPSGEKKDCPPCNPFGYNNGSSSCHPCPPGTFSDGT 420
Db 361 KTOIMYKWIIEPKICREDLTDAILRPPSGEKKDCPPCNPFGYNNGSSSCHPCPPGTFSDGT 420
Qy 421 KECRCPAGTEPALGFYKKNWVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIOGAGGS 480
Db 421 KECRCPAGTEPALGFYKKNWVLPGNMKTSCFNVGNSKCDGMNGWEVAGDHIOGAGGS 480
Qy 481 DNDYLILNLHLPGRKPPPTSMGTGATGSELGRITFVFETLCSADCVLVFMVDINRKSTNVVE 540
Db 481 DNDYLILNLHLPGRKPPPTSMGTGATGSELGRITFVFETLCSADCVLVFMVDINRKSTNVVE 540
Qy 541 SWGGTKEKOAYTHIIFKNATFTTFAFORTNOGQDNRFFINDMKIYISITATNAVGVAS 600
Db 541 SWGGTKEKOAYTHIIFKNATFTTFAFORTNOGQDNRFFINDMKIYISITATNAVGVAS 600
```

QY 601 SCRAALGSSQSSVCPPGHHYIEKETNOCKECPDYLSTHQVYKREACIPCGPGSK 660
PRIOR FILING DATE: 1997-09-24
PRIOR APPLICATION NUMBER: 60/062250
Db 601 SCRAALGSSQSSVCPPGHHYIEKETNOCKECPDYLSTHQVYKREACIPCGPGSK 660
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/062285
QY 661 NNQDHSVCSDCFFYHEKENQIILHYDFSNLSSVGLMNGPSETSKGTKYHFFFNISLCGH 720
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/062287
Db 661 NNQDHSVCSDCFFYHEKENQIILHYDFSNLSSVGLMNGPSETSKGTKYHFFFNISLCGH 720
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/062814
QY 721 EGKMACTNNITDFVVKELVAGSDDYTNLGVAFVCGSTIIPSEKGFRAALSSQSIILA 780
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/062816
Db 721 EGKMACTNNITDFVVKELVAGSDDYTNLGVAFVCGSTIIPSEKGFRAALSSQSIILA 780
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063045
QY 781 DTFIGVTETTLKINIKEDMPVPPTSQIPDVHFFKSTATTSCINGRSTAVKMCNPT 840
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/063127
Db 781 DTFIGVTETTLKINIKEDMPVPPTSQIPDVHFFKSTATTSCINGRSTAVKMCNPT 840
PRIOR FILING DATE: 1997-10-24
PRIOR APPLICATION NUMBER: 60/063327
QY 841 KSGAGVISVPSKCPAGTCDGCTFYFLWESAECPLCTEHDHFIEGACKRGFOETLYVWN 900
PRIOR FILING DATE: 1997-10-27
PRIOR APPLICATION NUMBER: 60/063329
Db 841 KSGAGVISVPSKCPAGTCDGCTFYFLWESAECPLCTEHDHFIEGACKRGFOETLYVWN 900
PRIOR FILING DATE: 1997-10-27
PRIOR APPLICATION NUMBER: 60/063350
QY 901 EPKWKIGISLPEKKLATCETVDFWLKVGAGVGAFTAVLLVALTCYFVKKNOK 953
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063561
Db 901 EPKWKIGISLPEKKLATCETVDFWLKVGAGVGAFTAVLLVALTCYFVKKNOK 953
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/063704
RESULT 3
US-10-028-072-38
Sequence 38, Application US/10028072
Publication No. US20030004311A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang
TITLE OF INVENTION:
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/10/028,072
CURRENT FILING DATE: 2001-12-19
PRIOR APPLICATION NUMBER: 60/049911
PRIOR FILING DATE: 1997-06-18
PRIOR APPLICATION NUMBER: 60/056974
PRIOR FILING DATE: 1997-08-26
PRIOR APPLICATION NUMBER: 60/059113
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059115
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059117
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059122
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059184
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/059263
PRIOR FILING DATE: 1997-09-18
PRIOR APPLICATION NUMBER: 60/059352
PRIOR FILING DATE: 1997-09-19
PRIOR APPLICATION NUMBER: 60/059588
PRIOR FILING DATE: 1997-09-19
PRIOR APPLICATION NUMBER: 60/059836


```
QY 940 LVALTCYFWKKKQKLEYKYKSKLVMTTNSKECELPAAADSCAIMEGEDNEEEVYVSNKOSLL 999
D 923 LTVTCYFWKKKQKLEYKYKSKLVMTTNSKECELPAAADSCAIMEGEDNEEEVYVSNKOSLL 982
QY 1000 GKLSLATKEKEDHESVOLKTS 1022
D 983 GKIKTSKRTPDGDSVPLKTS 1005

RESULT 4
US-10-121-049-38
; Sequence 38, Application US/10121049
; Publication No. US2003002239A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C17
; CURRENT APPLICATION NUMBER: US/10/121,049
; CURRENT FILING DATE: 2002-04-12
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 38
; LENGTH: 1013
; TYPE: PRT
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 877, 882
; OTHER INFORMATION: unknown amino acid
US-10-121-049-38
```

```
Query Match 53.2%; Score 3025; DB 9; Length 1013;
Best Local Similarity 54.4%; Pred. No. 1.3e-228;
Matches 535; Conservative 169; Mismatches 265; Indels 14; Gaps 9;

QY 46 WAG---DLPSSSRRLPPCQEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKE 102
D 31 WAGTAFQVQTGTGPELHACKSEIHYETACDGSRWVRVAPTPGLTSLSDPVKRGTE 90
QY 103 CTFSCASGEYLEMKQVCKSCGEGTYSLGSGIKFDWDELPAFGFSNIATFMDTVVGPDS 162
D 91 CSFSCNAGEFLDMQDSCKPCAEGRYSLGTGIRFDEWDELPHGFASLSANMELDDSAES 150
QY 163 RPDGNNNSWIRGRNIESNRDCTVLSIYAVHLKSGVVFVEYQYQVDMNIFEFFIQND 222
D 151 -TGNCSTSSKWVPRGDYIASNTDECTATLMYAVNLKQSGTVNFEYYPDSIIIFEFVQND. 209
QY 223 QCQEMDTTDDKVKLTLDNGEWSGVMKSGTNILYWRRTTGLMGSKAVKPVLYKNITIE 282
D 210 QCQP-NADSRNMKTEKG-WEFHSEVNLNRGNVLYWRRTAFSVMTKVPKPVLYRNIAIT 267
QY 283 GVAYTSECFPKPGTFSNKPGRNQCQVPRNTYSEKGAKEICRCKDDSOFS--GSSECTE 340
D 268 GVAYTSECFPKPGTVADKQSGSFCKLCPANSYKNGETSCHQCDPDKYSEKGSNCV 326
```

```
US-10-123-904-38
; Sequence 38, Application US/10123904
; Publication No. US20030022328A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME

RESULT 5
US-10-123-904-38
; Sequence 38, Application US/10123904
; Publication No. US20030022328A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
```

Tue May 13 13:15:34 2003

```
*
; FILE REFERENCE: P3330R1C54
; CURRENT APPLICATION NUMBER: US/10/123,904
; CURRENT FILING DATE: 2002-04-16
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 38
; LENGTH: 1013
; TYPE: PRT
; ORGANISM: Homo Sapien
; NAME/KEY: unsure
; LOCATION: 877, 882
; OTHER INFORMATION: unknown amino acid
US-10-123-904-38

Query Match 53.2%; Score 3025; DB 9; Length 1013;
Best Local Similarity 54.4%; Pred. No. 1.3e-228;
Matches 535; Conservative 169; Mismatches 265; Indels 14; Gaps 9;

Qy 46 WAG---DLPSSSRRLPPCQEKDYHFEYTCDSGSRWRVAIPNSAVDCSLDPPVRGKE 102
Db 31 WAGTAFQVQTGTGPPELHACKESYHYETACDSTGSRWRVAVPHPTGLCTSLSDPVKGT 90

Qy 103 CTFSCASGEYLEMKNQVCSKCGEGYSLGSGIKFDEWDELPAFESNIATFMDTVVGPDS 162
Db 91 CTFSCNAGEFLDMKDQCKPCAEGRYSGLTGIRFDEWDELPHGFASLANMELDDSAAES 150

Qy 163 RPDGNCNSWIPRGNYIESNRDCTVSLIYAVHLKSGVYFFEQYVDDNIFFEFFQND 222
Db 151 -TGNCITSSKWWPRGDYIASNTDECTATLMYAVNLKQSGTVNFEEYPPDSSIIFEFFQND 209

Qy 223 QCQEMDTTDDKWKLTUNGWGSVHLKSGTNILYWRRTGILMGSKAVKPVLVKNITIE 282
Db 210 QCQP-NADDSRWMMKTEKG-WEFHSELNNGNVLWRTTAFSVMTKPVPLVLRNIAIT 267

Qy 283 GVAYTSECPKPGTFNSKPGSFNCQVCPRTYSEKAKEICRCKDDSOFS--GSSECTE 340
Db 268 GVAYTSECPKPGTYADKQGSFCKLCPANSYKNGETSCHQC-DPDKYSEKSSSCNV 326

Qy 341 RPPCTTKDYFQITPCDEBEGTKIMYKWIETKICREDLDAIRLPPSGEKKDCPPCNPGF 400
Db 327 RPACTDKDYFTHACDANGETQIMYKWKAPKICSDELGAVKLPASGVKTHCPCNPGF 386

Qy 401 YNNGSSCHPCPGTSDGTEKCRPCPACTEPALGFYKWNVLPGNMTKSCFNVGNKSC 460
Db 387 FKTNNSTCQPCPYGYSNGS-DCTRCPCAGTEPAVGFEYKWNVLPTNMETTVLSGINFEY 445

Qy 461 DGNHGEVAGDHLOSAGGSDNDYLTLNLHIGFKPPTS-MTGATGSELGRITFEFETLC 519
Db 446 KGMTGEVAGDHIYTAAGASDNDFMILTLLVPGFRPPQVMADTENKEVARITFEFETLC 505

Qy 520 SADCVLVFMVDINRKSTNVVSMGGTKERQAYTHIIFKNATFTTWFQFRTNOGQDNRRF 579
Db 506 SVNCELYFVGVNSRTNTPVETWKGSKQSYIIEENTTTTFTWAFQRTTFHEASRY 565

Qy 580 INDMWYIYSTATNAVGVASSACALASEQSGSCVCPGPHYIETKQCKECPDPT 639
Db 566 TNDVAKIYSINVTNVANGVASYCRPCALEASDVGSCTSPAGYIIDRSGTCHSCPNT 625

Qy 640 YLSHQVYGEACIPCPGSKNODHVCYSDCFYHEKENOILHYDFSLSVSGSLMNG 699
Db 626 ILKAHQPYGVQACVPGCGPGTKNKHISLCYNDCTFSRNTPTRTFNFSALANTVTLAGG 685

Qy 700 PSFTSGTKYFHFENISLCHGEGKMACTNNITDTFTVKEIVAGSDDYTNLVGAFCQST 759
Db 686 PSFTSKGLKYFHFHTLSLCCNOKRKNVCTDNTDLRIPE--GESGFSKSITAYVCQAV 742

Qy 760 IIPESKGFRAALSSQSIILADFTIGVETTLKLNINIKEDMPVPTSQIPDVHFFKSS 819
Db 743 IIPPEVTGYKAGVSSQPVSLADRLIGVTTDMLDGTISPAELFHLESGLGIPDVIFFRSN 802

Qy 820 TATTSCINGRSTAVKMRNCTKSGAGVISVPSKCPACTCGCTFYFLWESAEACPLCTEH 879
Db 820 TATTSCINGRSTAVKMRNCTKSGAGVISVPSKCPACTCGCTFYFLWESAEACPLCTEH 879
```

```
QY 283 GVAYTSECFPCPKPGTFSNKPSENCQVCPNRTYSKGAKECIRCKDDSOFS--GSSECTE 340
|||||
Db 268 GVAYTSECFPCPKPGTADQKGSFCKLCPANSYSNKGTSCHQC--DPKYSKSGSSCNV 326
QY 341 RPPCTTKDYFQIHTPCDEBEGKTOIMYKWIPEKICREDLTDALRLPPSGEKKDCPCNPGF 400
|||||
Db 327 RPACTDKDYFYHTTACDANGETQLMYKNAKPKICSEDLGAVKLPASGVKTHCPCNPGF 386
QY 401 YNNGSSCHPCPGTFSNKPSENCQVCPNRTYSKGAKECIRCKDDSOFS--GSSECTE 460
|||||
Db 387 FKTNNSTCPCPGYSNKS--DCTRCPCAGTEPAVGFYKWNMTLPTNMTTSLGINFEY 445
QY 461 DGMNGWEVAGDHQSGAGSDNDYLILNLHIPGFKPPTS--MTGATGSELGRITVFETLC 519
|||||
Db 446 KGMTGWEVAGDHIYTAAGASDNDFMLTLVPGFRPPQSVMDTENKEVARITVFETLC 505
QY 520 SADCVLVFWMDINRSTNVNWSGGTKKQAYTHIIFKNATFTTFAFORTNOGQDNRRF 579
|||||
Db 506 SYNCELYFMVGVNSRTNTPVETWKGSKGQSYTYIEENTTTTFTWAFORTFHEASRKY 565
QY 580 INDMVKIYSITATNAVDGVASSCRACALGSESGSCVPCPPGHYIEKETNOCKECPDT 639
|||||
Db 566 TNDVAKIYSINVTNVMGVASVCRCALEASDVSGSCTSCPAGYIDRDSGTCHSCPPNT 625
QY 640 YLSIHQVYKKEACIPCGPGSKNNQDHSVCYDCFFYHEKENOILHYDFSNSLSSVGLMNG 699
|||||
Db 626 ILKAHQPYGVQACVPCPGTKNNKIHSLCYNDCTFSRNTPTPTFNYSALANTVTLAG 685
QY 700 PSFTSKGTGYHFFNLSLCGHEGKKMALCTNNITFTVKEIVAGSDDTNLVGFVQCST 759
|||||
Db 686 PSFTSKGLKYHFFHTLSLCGNOGRKMSVCTDNVTLRIPE--GESGFSKSIYAVVQAV 742
QY 760 IIPSESKGFRAALSSQSIILADTFIGTVETTLKNINIKEDMFPVPTSQIPDVHFFYKSS 819
|||||
Db 743 IIPPEVTGKAGVSQPSVSLADRLGVITDMLDITSPALFHLIESLGIPDVIFEYRSN 802
QY 820 TATTSCINGRSTAVKMRNPTKSGAGVISVPSKCPAGTCDGCTFYFLWESABACPLCTEH 879
|||||
Db 803 DVTQSCSSGRSTTIRVRCSPQKTPVPSLLLPCTGSDGTCDCGNFHLWESAACPLCSVA 862
QY 880 DPHEGACKRGFTQTLVYNWPKWICIGISLPEKKLATCEVDFWLKVGAGVGAFTAVL 939
|||||
Db 863 DYHAIYSSVAGIOXTTVXREPKLCSGGISLPEQRVICTIDFWLVKGISAGTCTAIL 922
QY 940 LVALTCYFWKNOKLEYKYSKLVMTTNSKECELPAADSCAIMEGEDNEEVVYSNKQSLL 999
|||||
Db 923 LTVLTCYFWKNOKLEYKYSKLVMTTNSKECELPAADSCAIMEGEDNEEVVYSNKQSLL 982
QY 1000 GKLSLAKTEREDHPESVOLKTS 1022
|||||
Db 983 GKIKSFTSKRTPDGFDSVPLKTS 1005
```

RESULT 7

US-10-175-746-38

Sequence 38, Application US/10175746

Publication No. US2003002720A1

GENERAL INFORMATION:

```
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Geritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
```

```
APPLICANT: Watanabe, Collin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C353
CURRENT APPLICATION NUMBER: US/10/175,746
CURRENT FILING DATE: 2002-06-19
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 38
LENGTH: 1013
TYPE: PRT
ORGANISM: Homo Sapien
NAME/KEY: unsure
LOCATION: 877, 882
OTHER INFORMATION: unknown amino acid
US-10-175-746-38
```

Query Match 53.2%; Score 3025; DB 9; Length 1013;

Best Local Similarity 54.4%; Pred. No. 1.3e-228;

Matches 535; Conservative 169; Mismatches 265; Indels 14; Gaps 9;

```
QY 46 WAG---DLPSSSRRLPPCOEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLDPVGRKE 102
|||||
Db 31 WAGTAFQVTOGTGPELHACKSEYHYEYACDSTCSRWRVAVPHTPGLCTSLSDPVKGT 90
QY 103 CTFSCASGEYLEMKNOVCSCGEGTYSLSGSIKFDWDELPAFSGNIATFMTDVTGVGPS 162
|||||
Db 91 CSFSCNAGFLDMQDQCKPCAEGYSLGTGRFDEWDELPHGFASLSANMELDDSAES 150
QY 163 RPDGNNSSWIPRGNYIESNRDDCTVSLIYVHLKKSQYVFEYQYVDNNIFEFFIQND 222
|||||
Db 151 -TGNCSTSSKWWPRGDIASNTDCTATLMYAVNLKQSGTVNFEYYPDSIIIEFEVQND 209
QY 223 QCQEMDTTDDKWKLTLDNGEMSGSHVLMKSGTNILYWRITGILMGSKAVKPLVKNITTE 282
|||||
Db 210 QCQP--NADDSRWKMTTEKG--WEFHSVELNRGNVLYWRITAFSVMTKVPKPLVNRNIAIT 267
QY 283 GVAYTSECFPCPKPGTFSNKPSENCQVCPNRTYSKGAKECIRCKDDSOFS--GSSECTE 340
|||||
Db 268 GVAYTSECFPCPKPGTADQKGSFCKLCPANSYSNKGTSCHQC--DPKYSKSGSSCNV 326
QY 341 RPPCTTKDYFQIHTPCDEBEGKTOIMYKWIPEKICREDLTDALRLPPSGEKKDCPCNPGF 400
|||||
Db 327 RPACTDKDYFYHTTACDANGETQLMYKNAKPKICSEDLGAVKLPASGVKTHCPCNPGF 386
QY 401 YNNGSSCHPCPGTFSNKPSENCQVCPNRTYSKGAKECIRCKDDSOFS--GSSECTE 460
|||||
Db 387 FKTNNSTCPCPGYSNKS--DCTRCPCAGTEPAVGFYKWNMTLPTNMTTSLGINFEY 445
QY 461 DGMNGWEVAGDHQSGAGSDNDYLILNLHIPGFKPPTS--MTGATGSELGRITVFETLC 519
|||||
Db 446 KGMTGWEVAGDHIYTAAGASDNDFMLTLVPGFRPPQSVMDTENKEVARITVFETLC 505
QY 520 SADCVLVFWMDINRSTNVNWSGGTKKQAYTHIIFKNATFTTFAFORTNOGQDNRRF 579
|||||
Db 506 SYNCELYFMVGVNSRTNTPVETWKGSKGQSYTYIEENTTTTFTWAFORTFHEASRKY 565
QY 580 INDMVKIYSITATNAVDGVASSCRACALGSESGSCVPCPPGHYIEKETNOCKECPDT 639
|||||
Db 566 TNDVAKIYSINVTNVMGVASVCRCALEASDVSGSCTSCPAGYIDRDSGTCHSCPPNT 625
QY 640 YLSIHQVYKKEACIPCGPGSKNNQDHSVCYDCFFYHEKENOILHYDFSNSLSSVGLMNG 699
|||||
Db 626 ILKAHQPYGVQACVPCPGTKNNKIHSLCYNDCTFSRNTPTPTFNYSALANTVTLAG 685
QY 700 PSFTSKGTGYHFFNLSLCGHEGKKMALCTNNITFTVKEIVAGSDDTNLVGFVQCST 759
|||||
Db 686 PSFTSKGLKYHFFHTLSLCGNOGRKMSVCTDNVTLRIPE--GESGFSKSIYAVVQAV 742
QY 760 IIPSESKGFRAALSSQSIILADTFIGTVETTLKNINIKEDMFPVPTSQIPDVHFFYKSS 819
```

```
Db 743 IIPPEVTGYKAGVSSQPSVSLADRLIGVTTDMTLDGTTSPAELFHLESUGIPDVIFYRSN 802
Qy 820 TATTSCINGRSTAVKMRNPTKSGAGVISVPSKCPAGTCDGCTFYFLWESAEACPLCTEH 879
Db 803 DVTQSCSSGRSTTIRVRCSPQKTVPGSLLPCTCDGCTDGCNFGHFLWESAACPLCSVA 862
Qy 880 DPHEIAGACKRGFTQYLLVYVNEPKWCIKIGISLPEKKLATCETVDFWLKVGAGVGAFTAVL 939
Db 863 DYHAIYSSCVAGIQXTYVYXREPCLKSCGSGISLPEQRVTICKTIDFWLKVGISAGTCTAIL 922
Qy 940 LVALTCYFWKKNQKLEYKYSKLVMTTNSKECELPAADSCAIMEGEDNEEVVYVSNKOSLL 999
Db 923 LTVLTCYFWKKNQKLEYKYSKLVMTTNSKECELPAADSCAIMEGEDNEEVVYVSNKOSLL 982
Qy 1000 GKLSKSLATKEEDHESVOLKTS 1022
Db 983 GKIKSFTSKRTPDGDSVPLKTS 1005

RESULT 8
US-10-176-918-38
; Sequence 38, Application US/10176918
; Publication No. US20030027275A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330RIC382
; CURRENT APPLICATION NUMBER: US/10/176, 918
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 38
; LENGTH: 1013
; TYPE: PRT
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 877, 882
; OTHER INFORMATION: unknown amino acid
US-10-176-918-38

Query Match 53.2%; Score 3025; DB 9; Length 1013;
Best Local Similarity 54.4%; Pred No. 1,3e-228;
Matches 535; Conservative 169; Mismatches 265; Indels 14; Gaps 9;

Qy 46 WAG---DLPSSSRPLPPQCKDYHFYEYTCDSGSRWRVAIPNSAVDCSGLPDPVRGKE 102
Db 31 WAGTAFQVQTGGPELHACKSEYHYEYACDSTGSRWRVAVPHTPGLCTSLSDPVKGTG 90
Qy 103 CTFSCASGEYLEMKVQCSKCEGTHYSLGKTFDEWDELPAFGFSNIATFMOTVVGPSDS 162
Db 91 CSFSCNAGFLDMKQDSCKPACBAEGRYSLGTGRFDEWDELPAFGFSANMELDOSAAES 150
Qy 163 RPDGCGNNSSWIPRGNVIESNRDDCTVSLIYAVHLKKGVVFFEQYQVDNNIFFERFIQND 222
```

```
Db 151 -TGNCSTSSKWWPRGDYIASNTDECTATLMYAVNLKQSGTVNFEYYPDSSIIFEFVQND 209
Qy 223 QCQEMDTTDDKWKVKLTDMNGEWSHMLKSGTNILYWRRTTGLMGSKAVKPVLVKNITIE 282
Db 210 QCOP-NADDSRWKMTKEG-WEFHSELNRGNVLVWRTTAFSVMTKVPKPLVRNIAIT 267
Qy 283 GVAYTSECPKPGTFSNKPQGSFNCQVCPNRTYSSKGAKECIRCKDDSOFS--GSSECTE 340
Db 268 GVAYTSECPKPGTAYADKQSGSFCKLCPANSYKNGETSCQOC-DPDKYSKSGSSCNV 326
Qy 341 RPPCTTDDYFOIHTPCDEBQKTOIMYKWTPEKICREDLTDALRLPPSGKKBKDCPPCNPQF 400
Db 327 RPACTDKDYFTHTACDANGETQLMYKNAKPKICSEDLGAVKLPASGVKTHCPCNPGF 386
Qy 401 YNNGSSCHPCPPGTFSDGTEKCRPCPACTGEPALGFYKMMNVLPNCNMKTSFCNVGNSKC 460
Db 387 FKTNNSTQCPQPGYSNGS-DCTRCPCACTEPAGVFEYKMMNTLPTNMTETVLVSGINFEY 445
Qy 461 DCMNGEAVAGDIIQSGAGSDNDYLILNLHIIPGRPPTS-WTGATGSELGRITTFEETLC 519
Db 446 KGMTGNEVAGDHIYTAAGASDNDFMILTLPVPCFRPQSVMAADTENKEYARTTFVPETLC 505
Qy 520 SADCVLYFMVDINRKSTNVVESWGGTKEQOAYTHIIFKNATFTTFAFORTNOGODNRFF 579
Db 506 SYNCELYFMVGVNSRINTPVETWKGSKGQSYTYIEENTTTSFTWAFORTFHEASRY 565
Qy 580 INDMVKIYSITATNAVGVASSCRACALGSESGSCVPCPPGHYIEKETNOCKECPPT 639
Db 566 TNDVAKIYSINVTVMNGVASYCRPCALEASDVSSCTSCPAGYVIDRDSGTHSCPPNT 625
Qy 640 YLSIHQVYGEACIPCGPGSKNNQDHSVCYDCFFYHEKENQILHYDFNSLSSVGLMNG 699
Db 626 ILKAHQPVGVQACVPCPGCTKNNKIHSCLYNDCTFSRNTPTRTFNYSALANTVTLAGG 685
Qy 700 PSFTSGKTYRHFENISLGGHEGKMKALCTNNTDFTVKELVAGSDDYTNLYCAFCVQST 759
Db 686 PSFTSGKLYFHHFTLSLGCNNGRMSVCTDNTDLRIPE---GESGFSKSTAYVYQNAV 742
Qy 760 IIPSESKGFRAALSSQSITLADTFICVTVETTLKNNIKEDMFPVPTSOIPDVHFYKSS 819
Db 743 IIPPEVTGYKAGVSSQPSVSLADRLIGVTTDMTLDGTTSPAELFHLESICIPDVIFYRSN 802
Qy 820 TATTSCINGRSTAVKMRNPTKSGAGVISVPSKCPAGTCDGCTFYFLWESAEACPLCTEH 879
Db 803 DVTQSCSSGRSTTIRVRCSPQKTVPGSLLPCTCDGCTDGCNFGHFLWESAACPLCSVA 862
Qy 880 DPHEIAGACKRGFTQYLLVYVNEPKWCIKIGISLPEKKLATCETVDFWLKVGAGVGAFTAVL 939
Db 863 DYHAIYSSCVAGIQXTYVYXREPCLKSCGSGISLPEQRVTICKTIDFWLKVGISAGTCTAIL 922
Qy 940 LVALTCYFWKKNQKLEYKYSKLVMTTNSKECELPAADSCAIMEGEDNEEVVYVSNKOSLL 999
Db 923 LTVLTCYFWKKNQKLEYKYSKLVMTTNSKECELPAADSCAIMEGEDNEEVVYVSNKOSLL 982
Qy 1000 GKLSKSLATKEEDHESVOLKTS 1022
Db 983 GKIKSFTSKRTPDGDSVPLKTS 1005

RESULT 9
US-10-176-921-38
; Sequence 38, Application US/10176921
; Publication No. US20030027276A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
```

APPLICANT: Gurney,Austin L.
APPLICANT: Sherwood,Steven
APPLICANT: Smith,Victoria
APPLICANT: Stewart,Timothy A.
APPLICANT: Tumas,Daniel
APPLICANT: Watanabe,Colin K
APPLICANT: Wood,William
APPLICANT: Zhang,Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C288
CURRENT APPLICATION NUMBER: US/10/176,921
CURRENT FILING DATE: 2002-06-20
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 38
LENGTH: 1013
TYPE: PRT
ORGANISM: Homo Sapien
FEATURE:
NAME/KEY: unsure
LOCATION: 877, 882
OTHER INFORMATION: unknown amino acid
US-10-176-921-38

Query Match 53.2%; Score 3025; DB 9; Length 1013;
Best Local Similarity 54.4%; Pred. No. 1.3e-228;
Matches 535; Conservative 169; Mismatches 265; Indels 14; Gaps 9;

QY 46 WAG---DLPSSSRPLPPCOEKDYHFEYTECDSSGSRWRVAIPNSAVDCSLPDPVRGKE 102
DB 31 WAGTAFVOTGTGPELHACKSEHYEYTCADSTGSRWRVAVPHPTGLCTSLSDPVKGTG 90
QY 103 CTFCASGEYLEMKNQVCSKCGEYSLGSGIKFDEWDELPAFESNIATFMDTVVGPDS 162
DB 91 CSFSCNAGEFLDMKDQSKCPKCAEGYSLGTCIRDEWDELPHGFASLSANWELDDSAES 150
QY 163 RPDGNNSSWIPRGNYTESRDDCTVSLIYAVHLKSGYFFQYVYDNNIFFEFFIOND 222
DB 151 -TGNCSTSKWVPRGDYIASNTDECTATLAVNLKQSGTVNFYYPDSLSIFEFFVQND 209
QY 223 QCQMDMTTDDKWLTDGWSHGVMLKSTNLLYVWRTTILGMSKAVKPVLVKNITIE 282
DB 210 QCQP-NADDSRWKTKEG-WEFHSVELNGNVLVWRTTAFSVWTKVPKPVLVNRNIAIT 267
QY 283 GVAYTSECFCKPGTFSNKPFSNCQVCPRTYSEKGAKECIRCKDSDQFS--GSSECTE 340
DB 268 GVAYTSECFCKPGTYADKQSSFCCLCPANSYSNKGTSCHQC-DPKYSEKSSSCNV 326
QY 341 RPPCTTKDYQIHTPCDEEGKTQIMYKWIPEKICREDLTAIRLPPSGEKKDCPPCNGF 400
DB 327 RPACTDKDYFTHACDANGETQIMYKWKAPKICSEDLGAVKLPAASGVKTHCPCPCNGF 386
QY 401 YNNGSSSCHPCPGTFSFGTKCPCPCAGTEPALGFFYKWNVLPGNMTKSCFNVNSKC 460
DB 387 FKTNNSTCQPCPGYSNGS-DCTRCPCAGTEPAVGFFYKWNVLPTNWTETVLGGINFEY 445
QY 461 DGMNGEVDAGDHQOSGAGSDNDYLIINLHIPGFKPPTS-MTATGSELGRITFVFETLC 519
DB 446 KGMTGWEVAGDHIYTAAGASDNDFMILTLLVVPGRPPQSVNADTENKEVARITFVFETLC 505
QY 520 SADCVLVEMVDIARKSNVSVESGCTKEKAYTHIIFKNATFTFTWAFQRTNOGQDNRRF 579
DB 506 SVNCELYFMVGVNSRTNPVETWKSFGKOSYTIIEENTTTSTFTWAFQRTTFTHEASRKY 565
QY 580 INDMVKIYSITATNAVGVASSACRALGSEQSGSSCCVCPGPHYIEKENTNOCKECPDPT 639
DB 566 TNDVAKIYSINVTNVMNGVASYCPCEALASDVGSSCTSPAGYIIDRDSGTCHSCPNT 625
QY 640 YLSTHQVYGEACIPCGPKGNQDHSCVYSDCFYFHEKENQILHDFSNLSSVGSILMNG 699
DB 626 ILRAHQPYGVQACVPCGPGTKNNKIHSCLYNDCTFSRNTPTRTFTFNFSALANTVTLAGG 685

QY 700 PSFTSKGTKYHFFENISLCGHEGKKMALCTNNITDTVKEIVAGSDDYTNLVGAFVCSOT 759
DB 686 PSFTSKGLKYHFFHTLSLCGNQGRKMSVCTDNVDLRPE---GESGFSKSIYAVCOAV 742
QY 760 IIPSEKGFRAALSSQSIIADTFIGVTVETTLKNINKEDMFVPVTSQIPDVHFFYKSS 819
DB 743 IIPPEVTGYKAGVSSQPVSLADRLTGVTDTMDLIGITSPAEFLHLESIGIPDVIFFYRSN 802
QY 820 TATTSCINGRSTAVKMRNCPKSGAGVISVPSKCPACTCDGCTFYFLWESAECPLCTEH 879
DB 803 DVTQSCSSGRSTTRVRCSPKTPVPGSLLLPCTGSDGTCDCGNHFLWESAACPLCSVA 862
QY 880 DFHEIEGACKRGFQETLYVWNEPKWIKIGISLPKIKLATCETVDFWLKVGAGVGAFTAVL 939
DB 863 DYHALVSSCVAGIQXTTYVXREPKLCSGGISLPQRTVICTIDFWLKVGLSAGTCTAIL 922
QY 940 LVALTCTFWKKNQKLEYKSKLVMTTNSKECELPAADESCAIMEGEDNEEVEVYVSNKQSL 999
DB 923 LTVLTCTYFWKKNQKLEYKSKLVMTNATLKCDLPAADESCAIMEGEDVEDDLFTSKKSLF 982
QY 1000 GKLSLATKKEKEDHFESVOLKTS 1022
DB 983 GKIKSFTSKRTPDGFDSVPLKTS 1005

RESULT 10

US-10-137-865-38
Sequence 38, Application US/10137865
Publication NO. US20030032155A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: DeForge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C154
CURRENT APPLICATION NUMBER: US/10/137,865
CURRENT FILING DATE: 2002-05-03
Prior Application removed - See Palm or File Wrapper
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 38
LENGTH: 1013
TYPE: PRT
ORGANISM: Homo Sapien
FEATURE:
NAME/KEY: unsure
LOCATION: 877, 882
OTHER INFORMATION: unknown amino acid
US-10-137-865-38

Query Match 53.2%; Score 3025; DB 9; Length 1013;
Best Local Similarity 54.4%; Pred. No. 1.3e-228;
Matches 535; Conservative 169; Mismatches 265; Indels 14; Gaps 9;

QY 46 WAG---DLPSSSRPLPPCOEKDYHFEYTECDSSGSRWRVAIPNSAVDCSLPDPVRGKE 102
DB 31 WAGTAFVOTGTGPELHACKSEHYEYTCADSTGSRWRVAVPHPTGLCTSLSDPVKGTG 90
QY 103 CTFCASGEYLEMKNQVCSKCGEYSLGSGIKFDEWDELPAFESNIATFMDTVVGPDS 162

Db 91 CSFSCNAGEFLDMKQDCKPCACGRYSGLTGIRDEWDELPHGFASLSANMELDDSAES 150
Qy 163 RPDGNNSSWIPRGNYSIESNRDDCTVSLIYAVHLKSGYVFEFYQYVNDNNTIFFEFTQND 222
Db 151 -TGNCSTSSKWPRGDIYASNTDECTATLMYAVNLKQSGTVNFEYYPDSIIFFEFVQND 209
Qy 223 QCOEMDTTDDKWLKLTONGEGSHSVMLKSGTNILYRTTGILMGSKAVKPVLVKNITIE 282
Db 210 QCQP-NADDSRWKMTTEKG-WEFHISVELNRNNVLYWRTAFSVMTKVPKPVLRNIAIT 267
Qy 283 GVAYTSECFPCPKPGTFSNKGPSFNCQVCPNRTYSEKGAKEICRCKDDSQFS--GSSECTE 340
Db 268 GVAYTSECFPCPKPGTYADKQGSFCKLCPANSYKNGKETSCHQC-DPDKYSEKSSSCNV 326
Qy 341 RPPCTTKDYFOIHTPCDEEGKTQIMYKWIPEKICREDLTDAILRPPSGEKKDCPCNPGF 400
Db 327 RPACTDKDYFTHTACDANGETQLMYKWKPKICSEDLGAVKLPASGVKTHCPNPGF 386
Qy 401 YNNGSSSCHPCPPGTFSDGTEKCRPCPAGTEPALGCFEYKMNVLPGNNKTSFCNVGNSKC 460
Db 387 FKTNNSTCQPCPYGSYNGS-DCTRCPAGTEPAGVGFYKKNWTLPTNNETTVLSGINFEY 445
Qy 461 DGMNGWEVAGDHIOGAGGSDNDYLLNLHITPGPKPTS-MTGATGSELGRITVFVETLC 519
Db 446 KGMTGWEVAGDHIYTAAGASDNDFMILTUVVPGFRPPQSVNADTENKEVARITVFETLC 505
Qy 520 SADCVLYPMVDINRKSNTNVESWGTEKQAYTHIIFKNATFTTWARFORTNOGQDNRRF 579
Db 506 SVNCELYPMVGVSNTNTPVETWKGSKGQSYTYIIEENTTSFTWARFORTFHEASRKY 565
Qy 580 INDVVKIYSITATNAVDGAVSSCRACALGSSQSSVCPCPGPHYIEKETNOCKECPDPT 639
Db 566 TNDVAKIYSINVTNVMGVASVCRPCALEASDVGSSCTSCPAGYIIDRDSGTCHSCPNT 625
Qy 640 YLSHQVYKGEACIPCGPGSKNODHVCYSDCFPYHEKENQIILHYDFSNLSSVGSLSMG 699
Db 626 ILKAHQPTGVQACVPCGPGTKNNKIHSLCYNDCTFSRNTPTRTFNYPFSALANTVTLAG 685
Qy 700 PSFTSKGTPYHFNISLUCHEGKMACTNNITDFTVKEIVAGSDDYTNLVGAVQOST 759
Db 686 PSFTSKGLKYFHHFTLSLCGNGRKMVSCTONVTDLRPE--GESGFSKSIYAVCOAV 742
Qy 760 IIPSEKGFRAALSQSILADTFIGVTVETTLKNINIKEDMFVPVTSQIPDVHFFYKSS 819
Db 743 IIPPEVTGKAGVSQPSVSLADRLIGVTTMDLDTGITSAPBLFLESGLIPDVIFFYRSN 802
Qy 820 TATTSCINGRSTAVKMRNPTKSGAGVLSVPSKCPAGTCDCGCTFYFLWESAEACPLCTEH 879
Db 803 DVTOSCSSGRSTTIRVRCSPQKTPVGSLLLPCTGSDGTCDCGCFHFLWESAAACPLCSVA 862
Qy 880 DFHETEGAKRGFBETLYVWNEPKWICKIGISLPKPKLATCETVDFWLKVGAGVGAFTAVL 939
Db 863 DYHALVSSCVAGIQXTTVXREPKLCSGGLSPLPQGVTVICKTIDFWLKVGISAGTCTAIL 922
Qy 940 LVALTCTYFWKNQKLEYKYSKLVMTNTSKEGELPAADSCALMBEDNEBEEVYVYKSKSL 999
Db 923 LTVLTCTYFWKNQKLEYKYSKLVMTNATLKDCLDPAADSCALMBEDNEBDDLIFTSKSLF 982
Qy 1000 GKLSKATKEKEDHESVOLKTS 1022
Db 983 GKIKSFTSKRTPDGFDSVPLKTS 1005

RESULT 11

US-10-140-474-38

; Sequence 38, Application US/10140474
; Publication No. US20030032156A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330RIG162
; CURRENT APPLICATION NUMBER: US/10/140,474
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See Palm or File wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 38
; LENGTH: 1013
; TYPE: PRT
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 877..882
; OTHER INFORMATION: unknown amino acid
; US-10-140-474-38

Query Match 53.2%; Score 3025; DB 9; Length 1013;
Best Local Similarity 54.4%; Pred. No. 1.3e-228;
Matches 535; Conservative 169; Mismatches 265; Indels 14; Gaps 9;
Qy 46 WAG---DLPSSSSRPLPCQEKDYHFEYTECDSSGRWRVAIPNSAVDCSLGPDVVRKE 102
Db 31 WAGTAFQQTGTGPELHACKSEYHYEYACDSTSGSRWRVAVPHPTPGTSLSDPDVKTE 90
Qy 103 CTFSACSEYLEMKNQVCSKCGEGTSLGSGIKFDEWDELPAFNSIATFMDTVVGPDS 162
Db 91 CSFSCNAGEFLDMKQDCKPCACGRYSGLTGIRDEWDELPHGFASLSANMELDDSAES 150
Qy 163 RPDGNNSSWIPRGNYSIESNRDDCTVSLIYAVHLKSGYVFEFYQYVNDNNTIFFEFTQND 222
Db 151 -TGNCSTSSKWPRGDIYASNTDECTATLMYAVNLKQSGTVNFEYYPDSIIFFEFVQND 209
Qy 223 QCOEMDTTDDKWLKLTONGEGSHSVMLKSGTNILYRTTGILMGSKAVKPVLVKNITIE 282
Db 210 QCQP-NADDSRWKMTTEKG-WEFHISVELNRNNVLYWRTAFSVMTKVPKPVLRNIAIT 267
Qy 283 GVAYTSECFPCPKPGTFSNKGPSFNCQVCPNRTYSEKGAKEICRCKDDSQFS--GSSECTE 340
Db 268 GVAYTSECFPCPKPGTYADKQGSFCKLCPANSYKNGKETSCHQC-DPDKYSEKSSSCNV 326
Qy 341 RPPCTTKDYFOIHTPCDEEGKTQIMYKWIPEKICREDLTDAILRPPSGEKKDCPCNPGF 400
Db 327 RPACTDKDYFTHTACDANGETQLMYKWKPKICSEDLGAVKLPASGVKTHCPNPGF 386
Qy 401 YNNGSSSCHPCPPGTFSDGTEKCRPCPAGTEPALGCFEYKMNVLPGNNKTSFCNVGNSKC 460
Db 387 FKTNNSTCQPCPYGSYNGS-DCTRCPAGTEPAGVGFYKKNWTLPTNNETTVLSGINFEY 445
Qy 461 DGMNGWEVAGDHIOGAGGSDNDYLLNLHITPGPKPTS-MTGATGSELGRITVFVETLC 519
Db 446 KGMTGWEVAGDHIYTAAGASDNDFMILTUVVPGFRPPQSVNADTENKEVARITVFETLC 505
Qy 520 SADCVLYPMVDINRKSNTNVESWGTEKQAYTHIIFKNATFTTWARFORTNOGQDNRRF 579
Db 506 SVNCELYPMVGVSNTNTPVETWKGSKGQSYTYIIEENTTSFTWARFORTFHEASRKY 565
Qy 580 INDVVKIYSITATNAVDGAVSSCRACALGSSQSSVCPCPGPHYIEKETNOCKECPDPT 639
Db 566 TNDVAKIYSINVTNVMGVASVCRPCALEASDVGSSCTSCPAGYIIDRDSGTCHSCPNT 625

QY 640 YLSIHQVYKACIPCGSKNQHDSVCYSDCFYHEKENOILHYDFSNLSSVGLMNG 699
 Db 626 ILKAHPYGVQACVPCGPKTKNKHSLCYNDCTFSRNTPTFTFYNFALANTVTLAG 685
 QY 700 PSFTSGKTYKFFHFNISLCGHEGKMACTNNITFTVKEIVAGSDDTNVLGAFVCO 759
 Db 686 PSFTSGKLYFHHFTLSLCGNGRMSVCTDNVTDLRIPE---GESGFSKSIYAVCO 742
 QY 760 IIPSEKGFRAALSQSIIADTFIGVTETLKNINIKEDMFPVPTSQIPDVHFFYKSS 819
 Db 743 IIPPEVTGYKAGVSQPVSLADRLIGVTTDLGITSPAEFLHLESIGIPVIFYSRN 802
 QY 820 TATTSINGRSTAVKMRNPTKSGAGVISVPSKCPAGTCDCGCTFYFLWESAEACPLCTEH 879
 Db 803 DVTQSCSSGRSTTIRVRCSPQKTVPGSLLLPCTCSDGTCDCGCFHLESAAACPLCSVA 862
 QY 880 DFHELEGACKRFOETLYVWNEPKWICIGISLPEKKLATCETVDFWLKVGAGVGAFTAVL 939
 Db 863 DYHAIVSSCVAGIQXTTYVXREPRLCSGISLPEQRVTICKTIDFWLKVGISAGTCTAIL 922
 QY 940 LVALTCYFWKKNQKLEYKYSKLVMTTNSKECELPAADESCAIMEGEDNEEVVYSNKOSLL 999
 Db 923 LTVLTCYFWKKNQKLEYKYSKLVMTATLKDCDLPAADESCAIMEGEDNEEDDLIFTSKSLF 982
 QY 1000 GKLSLATKEKEDHESVOLKTS 1022
 Db 983 GKIKSFTSKRTPDGFDSVPLKTS 1005

RESULT 12

US-10-142-431-38

; Sequence 38, Application US/10142431

; Publication No. US20030036179A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; FILE REFERENCE: P3330PLC251

; CURRENT APPLICATION NUMBER: US/10/142,431

; Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 38

; LENGTH: 1013

; TYPE: PRT

; ORGANISM: Homo Sapien

; NAME/KEY: unsure

; LOCATION: 877, 882

; OTHER INFORMATION: unknown amino acid

US-10-142-431-38

Query Match

Best Local Similarity 53.2%; Score 3025; DB 9; Length 1013;

Matches 535; Conservative 169; Mismatches 265; Indels 14; Gaps 9;

QY 46 WAG---DLPSSSRRLPPOCKDYHFEYTECDSSSRHRYVAIPNSAVDCSGLDPVPRCKE 102
 Db 31 WAGTAFQVOTGGPELHACKSEYHYETACDSTGSRWRVAVPHPTGLCTSLSDVPKGTG 90
 QY 103 CTFSCASGEYLEMKNQVCSKCGEGTYSLSGSIKGFDEWDELPAGFNSIATFMDTVVGP 162
 Db 91 CSFSCNAGEFLDMKQSCPKCAEGRYSLGTGIRFDEWDELPGFASLSANMELDDSA 150
 QY 163 RPDGNNNSWIPRGNYIESNRDDCTVLSIYAVHLKKSGYVFFEXYQYVDNNIFFE 222
 Db 151 -TGNTCKWPRGDIYIASNTDECTATIMYAVNLKQSGTVNEFYYPDSSIIFFE 209
 QY 223 OCOEMDDTTDKWKVLTDMGWSHVMKSGTNILYRTTGLMGSKAVKPVLVNITIE 282
 Db 210 QCOP-NADDSRMWKTTEKG-WEFHVELNRGNVLYWRTTAFVWTKVPKPVLVNIAIT 267
 QY 283 GVAYTSECFPCPKPGTFSNKGFCNQCVRNITYSKGAKECIRCKDDSOFS--GSSECTE 340
 Db 268 GVAYTSECFPCPKPGTYADKQSSFCCKLPANSYSNKGSETSCHQC-DPKYSEKSS 326
 QY 341 RPPCTTKDYFQIHTPCDEEGKTQIMYKWIPEKICREDLTDALRLPPSGEKKDCPC 400
 Db 327 RPACTDKDYFYHTTACDANGETQLMKWAKPKICSEDLEGAVKLPASGVKTHCP 386
 QY 401 YNNGSSSCHPCPGPTFSDGTEKRCPCPAGTEPALGFYKWNVLPGNMKTS CFNVGNSKC 460
 Db 387 FKTNNSTCQPCPYGSYNGS-DCTRCPAGTEPAVGFEYKWNWTLPTNMTETVLS 445
 QY 461 DGMNGWEVAGDHIOGAGGSDNDYLILNLHLPGEKPPTS-MTGATGSELGRITVF 519
 Db 446 KGMTGWEVAGDHIYTAAGASDNDPMILTLLVVPGRPPQSVMDATENKEVARIT 505
 QY 520 SADCVLYFWMVDINRKSNTNVSWGGTKEKQAYTHIIFKNATFTFTWAFORTNO 579
 Db 506 SVNCELXYFMVGNSRTPVETWKGSKQSYTYIIEENTTTTFTWAFORTTFEARS 565
 QY 580 INDMVKIYSITATNAVDCVASSCRACALGSESGSCVPCPGPHVIEKETNOCKECP 639
 Db 566 TNDVAKIYSINVTVMNGVASYCRPCALEASDVSGSCTSCPAGYIYDRDSTGCH 625
 QY 640 YLSIHQVYKACIPCGSKNQHDSVCYSDCFYHEKENQILHYDFSNLSSVGLMNG 699
 Db 626 ILKAHPYGVQACVPCGPKTKNKHSLCYNDCTFSRNTPTFTFYNFALANTVTL 685
 QY 700 PSFTSGKTYKFFHFNISLCGHEGKMACTNNITFTVKEIVAGSDDTNVLGAFVCO 759
 Db 686 PSFTSGKLYFHHFTLSLCGNGRMSVCTDNVTDLRIPE---GESGFSKSIYAVCO 742
 QY 760 IIPSEKGFRAALSQSIIADTFIGVTETLKNINIKEDMFPVPTSQIPDVHFFYKSS 819
 Db 743 IIPPEVTGYKAGVSQPVSLADRLIGVTTDLGITSPAEFLHLESIGIPVIFYSRN 802
 QY 820 TATTSINGRSTAVKMRNPTKSGAGVISVPSKCPAGTCDCGCTFYFLWESAEACPLCTEH 879
 Db 803 DVTQSCSSGRSTTIRVRCSPQKTVPGSLLLPCTCSDGTCDCGCFHLESAAACPLCSVA 862
 QY 880 DFHELEGACKRFOETLYVWNEPKWICIGISLPEKKLATCETVDFWLKVGAGVGAFTAVL 939
 Db 863 DYHAIVSSCVAGIQXTTYVXREPRLCSGISLPEQRVTICKTIDFWLKVGISAGTCTAIL 922
 QY 940 LVALTCYFWKKNQKLEYKYSKLVMTTNSKECELPAADESCAIMEGEDNEEVVYSNKOSLL 999
 Db 923 LTVLTCYFWKKNQKLEYKYSKLVMTATLKDCDLPAADESCAIMEGEDNEEDDLIFTSKSLF 982
 QY 1000 GKLSLATKEKEDHESVOLKTS 1022
 Db 983 GKIKSFTSKRTPDGFDSVPLKTS 1005

RESULT 13

US-10-143-114-38

; Sequence 38, Application US/10143114

; Publication No. US20030036180A1

Query Match 53.2%; Score 3025; DB 9; Length 1013;
Best Local Similarity 54.4%; Pred. No. 1.3e-228;
Matches 535; Conservative 169; Mismatches 265; Indels 14; Gaps 9;

Qy 46 WAG---DLPSSSRRLPPOCKDYHEFYTECDSSSRWRVAIPNSAVDCSLDPVGRKE 102
Db 31 WAGTAFQVQTGGPELHACKSEYHYEYACDSTGSRWRVAVPHTPGLCTSLSDPVKQTE 90
Qy 103 CTFSCASGEYLEMKNQVCKCEGYSLGSGIKFDEWDELPAFESNIATFMDTVVGPDS 162
Db 91 CSFSCNAGEFLDMKQSCPKCAEGRYSLGTGIRFDEWDELPHGFASLSANMELDDSAES 150
Qy 163 RPDGNNSSWIPRGNYIESNRDDCTVSLIYAVHLKSGYVFEYVDNNIIEFFEQND 222
Db 151 -TGNCSTSSKWPGRDGIASNTDECTATLMYAVNLKQSGTVNFEYYPDSIIIEFFVQND 209
Qy 223 OCQEMDTTDDKWLTDNGEWSHVMKSGTNILYWRRTGILMGSKAVKPVLVKNITIE 282
Db 210 OCQP-NADDSRMKTKTEG-WEFHVELNRGNVLYWRTAFSVMTKPKVPLVRNIAIT 267
Qy 283 GVAYTSECFPCPKPGTFSNKPGSFNCQVCPNTYSEKGAKEICRCKDDSOFS--GSSECTE 340
Db 268 GVAYTSECFPCPKPGTYADKQGSFCKLCPANSYKNGETSHQC-DPKYSEKSSSCNV 326
Qy 341 RPPCTTKDYFQIHTPCDEEGKTQIMYKWIPEKICREDLTDAIRLPPSGEKKDCPCNPGF 400
Db 327 RPACTDKDYFTHYACDANGETQLMYKWKAPKICSEDLGAVKLPASGVKTHCPCNPGF 386
Qy 401 YNNGSSSCHPCPPGTFSDGTEKRCPCPAGTEPALGFEYKWNVLPGNMKTSCFNVGNKSC 460
Db 387 FKTNNSTQCPQCPYSGYSNGS-DCTRCPAGTEPAVGFEYKWNVLTPTNMETTVLSGINFEY 445
Qy 461 DGMNGWEVAGDHIOGAGGSDNDYLILNLHIPGPKPPTS-MTGATGSELGRTTFVFETLC 519
Db 446 KGMTGWEVAGDHIYTAAGASDNDFILVLVPGFRPPQSVMADETENKEVARITTFVFETLC 505
Qy 520 SADVLYFMDINRKNVSVESGWTKEQAYTHIIFKNATFTFWAFORTNQGDNRFF 579
Db 506 SYNCLELYFMVGNSTNPTVETWKGSKQSYTIEENTTTTFTWAFORTTFEASRKY 565
Qy 580 INDMYKISITATNAVDGVASSCRALGSEQSSVCPCPPGHVIEKTNCKECPDPT 639
Db 566 TNDVAKIYSINVTNNGVNVASVCRPCALGSEADVGSSCTSCPAGYIDRDSGTCHSCPNT 625
Qy 640 YLSIHQVYKCAICPCPGSKNODHVCYSDFEYHEKENOILHYDFSNLSSVSGSLMNG 699
Db 626 ILKAHQPYGVQACVPCGPKTKNKHLSYNDCTFSRNTPTPTFNYSALANTVTLAGG 685
Qy 700 PSFTSKGTKYFHFNLSCGHEGKMAKCTNNITDFTVKEIVAGSDDTNLVGAFCVQST 759
Db 686 PSFTSKGLKYFHFHTLSLGNQGRKMSYCTDNVTDLRIPE---GESGFSKSIYAVCQAV 742
Qy 760 IPPSEKGFRAALSQSILADTFGTGIVETTLKNINIKEDMFVPTSOIPDVHFFYKSS 819
Db 743 IIPPEVTYKAGVSQPVSLADRLIGVTDMLTGITSPAEFLHESLGIPDVIFVFSN 802
Qy 820 TATTSCINGRTAVKMRNPTKSGAGVISVPSKCPAGTCDCGCTFFFLWESAEACPLCTEH 879
Db 803 DVTQSCSSGRSTTIRVRCSPQKTPVPSLLLPCTCSGDCGDCGNFHLWESAAACPLCSVA 862
Qy 880 DFHETEGACKRGFQETLYWNEPKWCIKIGSLPEKKLATCTETVDFWLKVGAGVGAFTAVL 939
Db 863 DYHALVSSCVAGIOXTYVXREPKLCSGIGSLPEQORVITCKTIDFWLVGVSAGTCTAIL 922
Qy 940 LVALTCYFWKNOKLEYKYKSLVMTNTSKECELPAADESCAIMEGEDNEEVVYVSNQSL 999
Db 923 LTVLTCYFWKNOKLEYKYKSLVMTNATLKDCLPAADSCAIMEGEDVDDLIFTSKSLF 982
Qy 1000 GKLKSLATKEKEDHESVOLKTS 1022
Db 983 GKIKSFTSKRTPDGDSVPLKTS 1005

RESULT 15
US-10-142-419-38
; Sequence 38, Application US/10142419
; Publication No. US20030044945A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C244
; CURRENT APPLICATION NUMBER: US/10/142,419
; CURRENT FILING DATE: 2002-05-10
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 38
; LENGTH: 1013
; TYPE: PRT
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 877, 882
; OTHER INFORMATION: unknown amino acid
US-10-142-419-38

Query Match 53.2%; Score 3025; DB 9; Length 1013;
Best Local Similarity 54.4%; Pred. No. 1.3e-228;
Matches 535; Conservative 169; Mismatches 265; Indels 14; Gaps 9;

Qy 46 WAG---DLPSSSRRLPPOCKDYHEFYTECDSSSRWRVAIPNSAVDCSLDPVGRKE 102
Db 31 WAGTAFQVQTGGPELHACKSEYHYEYACDSTGSRWRVAVPHTPGLCTSLSDPVKQTE 90
Qy 103 CTFSCASGEYLEMKNQVCKCEGYSLGSGIKFDEWDELPAFESNIATFMDTVVGPDS 162
Db 91 CSFSCNAGEFLDMKQSCPKCAEGRYSLGTGIRFDEWDELPHGFASLSANMELDDSAES 150
Qy 163 RPDGNNSSWIPRGNYIESNRDDCTVSLIYAVHLKSGYVFEYVDNNIIEFFEQND 222
Db 151 -TGNCSTSSKWPGRDGIASNTDECTATLMYAVNLKQSGTVNFEYYPDSIIIEFFVQND 209
Qy 223 OCQEMDTTDDKWLTDNGEWSHVMKSGTNILYWRRTGILMGSKAVKPVLVKNITIE 282
Db 210 OCQP-NADDSRMKTKTEG-WEFHVELNRGNVLYWRTAFSVMTKPKVPLVRNIAIT 267
Qy 283 GVAYTSECFPCPKPGTFSNKPGSFNCQVCPNTYSEKGAKEICRCKDDSOFS--GSSECTE 340
Db 268 GVAYTSECFPCPKPGTYADKQGSFCKLCPANSYKNGETSHQC-DPKYSEKSSSCNV 326
Qy 341 RPPCTTKDYFQIHTPCDEEGKTQIMYKWIPEKICREDLTDAIRLPPSGEKKDCPCNPGF 400
Db 327 RPACTDKDYFTHYACDANGETQLMYKWKAPKICSEDLGAVKLPASGVKTHCPCNPGF 386
Qy 401 YNNGSSSCHPCPPGTFSDGTEKRCPCPAGTEPALGFEYKWNVLPGNMKTSCFNVGNKSC 460
Db 387 FKTNNSTQCPQCPYSGYSNGS-DCTRCPAGTEPAVGFEYKWNVLTPTNMETTVLSGINFEY 445
Qy 461 DGMNGWEVAGDHIOGAGGSDNDYLILNLHIPGPKPPTS-MTGATGSELGRTTFVFETLC 519

Db 446 KGMTGWEVAGDHIYTAAGASDNDFMILTVVPGRPQSVMDTENKEVARITFVFETLC 505
Qy 520 SADCVLYEMVDINRKSTNVVSWGTEKEQAYTHIIFKNATFTTWARFORTNOGDNRRF 579
Db 506 SVNCLYFMVGVNSRNTNPVETWKGSKGQSTYIIIEENTTSFTWARFORTTFFEASRKY 565
Qy 580 INDMVKIYSITATNAVDGAVSSACALGSEOSGSCVPCPPGHYIEKETNOCKECPDPT 639
Db 566 TNDVAKIYSINVTNMVNGVASYCRPCALEASDVGSSTSCPAGYYIDRDSGTCHSCPNT 625
Qy 640 YLSIHQVYKGEACTPCGPGSKNNQDHVCYSDCFYHEKENQIILHYDFSNLSSVGSMLNG 699
Db 626 ILKAHQPYGVQACVPCGPGTKNNKTHSLCYNDCTFSRNTPTRTENYFNSALANTVTLAGG 685
Qy 700 PSFTSKGTKYHFFHFNISLCGHEGKKMALCTNNITDFTVKEIVAGSDDYTNLVGAFVCOST 759
Db 686 PSFTSKGLKYHFFHFTLSLCGNQGRKMSVCTDNVTDLRPE---GESGFSKSIYVCQAV 742
Qy 760 IIPSESKGFRAALSSQSIIADTFIGVTVEITLKNINIKEDMFPVPTSQIPDVHFFYKSS 819
Db 743 IIPPEVTGYKAGVSSQPVSLADRLIGVTTDMTLDGITSPAELFHLESIGIPDVIFFYRSN 802
Qy 820 TATTSCINGRSTAVKMRNPTKSGAGVISVPSKCPAGTCDGCTFYFLWESABACPLCTEH 879
Db 803 DVTQSCSSGRSTTIIRVCGSPQKTVPGSLLLPCTGSDGTCDCGNFHLWESAAACPLCSVA 862
Qy 880 DFHEIEGACKRGFOETLYVMNEPKVICIGISLPEKKLATCETVDFWLKVAGVGAVTAVL 939
Db 863 DYHAIVSSCVAGIQXTTVYXREPDKLCSGISLPEORVITCKTIDFWLVKVISAGTCTAIL 922
Qy 940 LVALTCYFWKKKQLEYKYSKLVMTTNSKECELPAAADSCAIMEGEDNEEEVYYSNKQSLL 999
Db 923 LTVLTCYFWKKKQLEYKYSKLVMNATLKCDDLPAADSCAIMEGEDVEDDLIFTSKKSILF 982
Qy 1000 GKLSLATKEKEDHESVOLKTS 1022
Db 983 GKIKSFTSKRTPDGFDSVPLKTS 1005

Search completed: May 12, 2003, 13:37:33
Job time : 191.369 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 11:39:43 ; Search time 29.9327 Seconds
(without alignments)
3298.405 Million cell updates/sec

Title: US-10-073-333A-4
Perfect score: 5681
Sequence: 1 MLFRAGPVRGRGWRPAEA.....KEKEDHFESVOLKTRSPNI 1027

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73.*

- 1: pir1.*
- 2: pir2.*
- 3: pir3.*
- 4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	242.5	4.3	1548	2 S34583	serine proteinase
2	210	3.7	713	2 A35502	major surface-labe
3	200	3.5	3635	2 T10053	laminin alpha 5 ch
4	199.5	3.5	2823	2 T23084	hypothetical prote
5	199.5	3.5	2823	2 F87908	protein T22A3.8 [i
6	199.5	3.5	3102	2 T43291	laminin alpha chai
7	198.5	3.5	677	2 C42125	trophozoite cystei
8	194	3.4	5376	2 T42215	zonadhesin - mouse
9	186	3.3	1766	2 A42125	trophozoite cystei
10	183	3.2	1372	2 T25919	hypothetical prote
11	182	3.2	667	2 A48579	hypothetical prote
12	181	3.2	1790	1 MMFFB1	trophozoite surfac
13	178.5	3.1	3672	2 T23433	laminin beta-1 cha
14	178.5	3.1	3704	2 T37316	hypothetical prote
15	177.5	3.1	1607	1 MMMSB2	probable laminin a
16	177.5	3.1	2907	2 A52728	laminin gamma-1 ch
17	174.5	3.1	596	2 A45664	fibrillin-2 precu
18	173.5	3.1	2871	2 A55624	variant-specific s
19	173	3.0	1111	2 T26972	fibrillin-1 precu
20	173	3.0	1786	1 MMMSB1	hypothetical prote
21	171	3.0	1620	2 T27283	laminin beta-1 cha
22	170.5	3.0	2437	2 S42612	hypothetical prote
23	170	3.0	1557	2 T28811	transmembrane prot
24	170	3.0	1786	1 MMHUB1	hypothetical prote
25	166.5	2.9	557	2 A48434	laminin beta-1 cha
26	166.5	2.9	1274	2 T42017	variant-specific s
27	166	2.9	3002	2 A47221	cysteine rich prot
28	165	2.9	1639	1 MMFFB2	fibrillin 1 precu
29	164.5	2.9	2918	2 A54105	laminin gamma-1 ch
					fibrillin-2 precu

30	164	2.9	2491	1 A28372	insulin-like growt
31	164	2.9	2871	2 A55567	fibrillin 1 - bovi
32	164	2.9	3084	1 MMMSA	laminin alpha-1 ch
33	163	2.9	1297	2 T30274	proteoliasin - se
34	161	2.8	3712	2 S18253	laminin alpha-1 ch
35	160.5	2.8	1609	1 MMHUB2	laminin gamma-1 ch
36	160	2.8	1220	2 A56136	jagged protein pre
37	159.5	2.8	2352	2 T30201	Notch homolog prot
38	157	2.8	2531	2 A46019	Notch-1 protein -
39	156.5	2.8	3106	1 S38688	laminin alpha-2 ch
40	155	2.7	2499	1 A30788	mannose 6-phosphat
41	154.5	2.7	13288	2 T03099	mucin, submaxillar
42	154	2.7	1323	2 P00568	connectin 3B - chi
43	153.5	2.7	1680	2 A43434	furin (EC 3.4.21.7
44	153	2.7	1959	1 AGRT	agrin - rat
45	153	2.7	2318	2 S45306	notch 3 protein -

ALIGNMENTS

RESULT 1

S34583
serine proteinase (EC 3.4.21.-) PC6B - mouse
C:Species: Mus musculus (house mouse)
C:Date: 02-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 05-Nov-1999
C:Accession: S34583
R:Nakagawa, T.; Murakami, K.; Nakayama, K.
FEBS Lett. 327, 165-171, 1993
A:Title: Identification of an isoform with an extremely large Cys-rich region of PC6,
A:Reference number: S34583; MUID:93327934; PMID:8335106
A:Accession: S34583
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1548 <NAK>
A:Cross-references: GB:D17583; NID:g407344; PIDN:BAA04507.1; PID:d1005033; PID:g44037
C:Keywords: hydrolase; serine proteinase

Query Match 4.3%; Score 242.5; DB 2; Length 1548;
Best Local Similarity 18.4%; Pred. No. 5.9e-08;
Matches 248; Conservative 116; Mismatches 432; Indels 549; Gaps 72;

QY	61	COEKDYHFEYTECDSSGSRWRVAIPNSAVDCSGLPDPVRGKEC-----TFSCA 108
DB	347	CPPGHYHADKKRC-----RKAPN-CESCFG-----SHGDOCLSKCYGFLEETSSCV 394
QY	109	S-----GEYLEMKNQVCKGEGTYSLGSGIKFDEWDELPAFES-----NIATFMDTVVG 158
DB	395	TQCPDGSYEDIKKNVCGKCSNCKAC---IGPHNCTECKGLSLQSGSCSVTCEDQGQFEN 451
QY	159	PSDSRP-----DCNNSWIPRGNYIESNR--DDCTVSLIYAVHLKSGYVF 203
DB	452	GHDQCPCHRFATCSGAGAGDCINCT---EGYVMEGRVCQSCSVS--YLDHSSEGGYK- 506
QY	204	FEYQYVDNNIFF-----EFFIQNDQCQEMDTT 231
DB	507	-SCKRCDNSCLTCNGPFGKNCSSCPGYLLDLGTGCOMGAICKGEYIDDOGHQCTCEASC 565
QY	232	DKWVKLTONGENGSHVLMKSGTNILYWRRTGILNGSKAVKPVLYKNITIEGAVYTSECF 291
DB	566	AKC-----WG-----PTQEDCISCPVTRVLDDGRC-----VMNCPMSWKFEEKQCH 606
QY	292	PCKPOTFSNKGFSNFCVCPNRTYSEKAKCEIRCKDD-----SQFSGSSECTERPCPTTK 347
DB	607	PC-----HYTCQGC-----OGSGPSNCTSCRADKHGOERELYHGECLN---CPVG 649
QY	348	DY-FOIHT--PCDESGKTQIMYKWIPEKICREDLTDALRLPPS-----GE 389
DB	650	HYPAGHGTCLPCPD--NCELCY---NPHICRSMGSGYLIIPNHTCQKLECRQGFQDSE 704
QY	390	KKDCPPCPNGFY---NNGSSSCHPCPPGTFSDGTEKCRPCAGTEPALGFYKKNVLPFG 446
DB	705	YEECPCEEGCLGCTEDDGGACTSCATGYMFERHCYKACPEKT-----FGVKWECRACG 759

QY 447 NMKTSFNVGNSKCDGMNGWEVAGDHQISGAGGSDNDYLILNLHIPPGRKPTMTGATGS 506
DB 760 ---TNGSCDOHEC---YWCSEGFELSGSCVQDCG---PGFH-----GDO 796
QY 507 ELGRITFVFETLCSADCVLVYFMDINRST--NVVSENGGTKEKQAYTHIFKNATFTFT 564
DB 797 ELGEC-----KPCRACETCTGSGYNOCSSCOEGLQHLWHTCLMWTWPQVEKGD----- 845
QY 565 WAFORTNQG-----QDNRREINDMVKIYSITATNAVGVASSCRACALGSEQ 611
DB 846 W-----NEAVPTKPSLVRSLLODRKW---KVQI-KRDATSONQCHSSCKTC----- 890
QY 612 SGSSCVPCPPHYI-----ERETNOCKEC-----PPD 638
DB 891 NGLSCLASCTPGMYLWLQACVPCPOGTWPSVTSGCCSEKSDCVSCSGADLCOQCCLSQPD 950
QY 639 TYLSIHQ-----VYKKE-ACIPCGPGSKNQ-----DHSVCY--- 669
DB 951 NTLLHLEGRYHSCPEGFYAKDGVCEHCSSPCKTCGNATSCNCEGDFVLDHGVCMWKC 1010
QY 670 -----SDCFYHEK-----ENQILHYDFSLSVGSGL----- 696
DB 1011 PEKHVAVGVCHKPERODCI--HEKTKCEMPDFFLYNDMCHRSCPKSFYPMRQCV 1068
QY 697 -----MNGPSFTSGTKYFHFNISLGHGKKM--ALCTNNITDFTVKE----- 739
DB 1069 CHKNLECGPKED-----DCKVCADTSKALHNGLCLECEPGYKEEENDECD 1118
QY 740 -----IVAGSDDTNLV-----GAFVCGSTIIPSESKEGR 769
DB 1119 CPESCLICSSAWTCLACRGFTVHDVCTAPKEAAVEYWDGSHRCQ-----PCHKCSR 1174
QY 770 AALSQSIIADTFICGVTVETTLKIN-ITKEDMFVPTSQIPDVHFFYKSTATTSC--- 825
DB 1175 CSGPSE-----DQYTCPRETELNTTCVKE-----CPEGYHTDKDSQCVLCHSS 1220
QY 826 ---INGRSTAVRMNCPKTSAGVISVPSKC-----PAGTCDCG----- 861
DB 1221 CRTCEGPHSMOCLSCR-----GWOLFGECLLQCRDGYGEGSTSGRCEKDKSCKSCRG 1275
QY 862 -----THYFLWESAEAC-PLCTEHDF-----HEGAC--KRGQETLY 897
DB 1276 PRPTDCQCDTFFLLRSKGQCHRAQPEHYADQHAQTCERCPTCDKCSGKEAMSLCSC 1335
QY 898 VNEPKWCIKGISLPE-----KKLATCET----- 921
DB 1336 VMS--YHLLKGICIPCIVGEYREGKGFENCKKCHESCMCKGPGSKNCTCSAGLLLD 1393
QY 922 -----VDFW-----LKVAGVGAFATV 938
DB 1394 MDDNRLHCCNASHRRSODCCQSDSTDECILPAREAFYEHTKALLVTSGA---MLL 1450
QY 939 LLVALTCYFNKNQK---LEYKSKLVMTNS-KECELPAADESCAIME-----GEDNEE 988
DB 1451 LLGAAVWNRKSRSPRAKRYEKLAEPYTSYSSRSYSLDEQDQVIEYRDYDEDDDED 1510
QY 989 EVVYSNKO-SLLCKKLSLATKEKED 1012
DB 1511 DIVYMGQGTVYRKFKYGLLDETED 1535

RESULT 2
A35502
major surface-labeled trophozoite antigen precursor - Giardia lamblia
C:Species: Giardia lamblia
C:Date: 09-Nov-1990 #sequence_revision 09-Nov-1990 #text_change 09-Sep-1997
R:Gilllin, F.D.; Hagblom, P.; Harwood, J.; Aley, S.B.; Reiner, D.S.; McCaffery, M.; So, M.
Proc. Natl. Acad. Sci. U.S.A. 87, 4463-4467, 1990
A:Title: Isolation and expression of the gene for a major surface protein of Giardia lam
A:Reference number: A35502; MUID:90280395; PMID:2352929
A:Accession: A35502

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-713 <G1>
A:Cross-references: GB:M33641; NID:g159131; PID:g159132
C:Keywords: surface antigen; transmembrane protein

Query Match 3.7%; Score 210; DB 2; Length 713;
Best Local Similarity 19.7%; Pred. No. 3.4e-06;
Matches 186; Conservative 96; Mismatches 300; Indels 360; Gaps 53;

QY 90 DCSGLP-DPVRCCKECTFSCASGEYLEMKNQ---VCSKCG-----EGTYSLSGSGIK-- 135
DB 43 ECNGANYAPVNG-QCDVNAEGPSKTLCPQHISAGKCTCCGGNSFMYKDCGSSYSGGLPGH 101
QY 136 -----FDEWDEL-----PAGFSN-----IATEMDTVVGPSPD---S 162
DB 102 SLICLSSDGGVCTEAAAPGYFAPVGAANTEQSVIACGDTTGVIIAAGNTYKGIADCAECS 161
QY 163 RPDGNNSSWIPRGNVIESNRDDCTVSLIYAVHLKKSQVYFVEYQYVDNNIIFEFQND 222
DB 162 APDA---TAGAAGKAVATCTK--CGVS-----KYLKDNVCV-----D 193
QY 223 QCOEMDTTIDKWKVLTDCNGEMGSHSVMLKSGTNILYWRITGILMGSKAVKPVLVKNITIE 282
DB 194 KAQCNGSGSTNKEVAV--DDESENGKCVSCSDNLN-----G 226
QY 283 GVAYTSECFPCPKGTFSNKPQSGFNCQVCPRTNYSEKGAKECIRCKDDPSQSGSSE--C 338
DB 227 GVA-----NCDTCA---SYDQSKKIKCTCTDNNYLKTTSEGTSC 263
QY 339 TERPCTTKDYFQIHTPDEEKTQIMYKWIIEPKITCREDLTDAILRPPSGEKKDPPCPNP 398
DB 264 VQKQDC--KQGF--FPKDDSS-----AGNK--CLIPCND 290
QY 399 GFYNNSSSCHPCP--PGTFSGDTKECRPCPACTEPALGFYKWNVLPGNMTSCFNVG 456
DB 291 S--TDGIANCATCALVSGRSGAALVTSACTDGYKPSAD-----KTTCEAVS 335
QY 457 NSKCDGMNGWEVAGDHQISGAGGSDNDYLILNLHIPPGRKPTMTGATGSELGRITTFVPE 516
DB 336 NCKTPCCRACSNEGKENEVCTDCDGSYLV-----TPTSQ----- 369
QY 517 TLCSADCVLVYFMDINRSTNVVSENGGTKEKQAYTHIFKNATFTFWAFORT--NQOQ 574
DB 370 --CIDSQA-----KIGNYVYATEGAKK-----LCKECTA---ANCKTCDQDQ 407
QY 575 DNRREINDMVKIYSITATNAVGVASSCRACALGSEOSGSSCVCPGPHYIEK-----ET 629
DB 408 -----COACNDGFKNGDACSPC---HESCKTCSAGTA 437
QY 630 NOCKECPDPTYLSIHQVYKKEACI-PCGPGSKNNQDHSVYSDCFYHEKENQILHYDFS 688
DB 438 SDCTECPCTKALR---YGDGDTKGTGCGGCTGTGAGAC----- 473
QY 689 NLSSVGLMNGSFSTSGTKYFHFENISLCGHEGKKMALCTNNITDFTVKEIVAG--SD 746
DB 474 --KTCLGLTIDGASCYSECATTTTEYPQNGVC---APKASRATPTCNDSPIONGVCGTCADN 528
QY 747 YTNLVCAFVCQSTIIPSESKGFRAALSOSIILADTFICGVTVETTLKNNIKEDMFVPT 806
DB 529 YFKMNGG--CYETV---KYPGKTVCLISAPN-----GGTCOKAANDGYKLDGSLTIVCS 575
QY 807 SQIPDVHFFYKSTATTSCING--RSTAVKMRN---PTKSGAGVTSVPKCPAGTCDCG 861
DB 576 EGCKEC---ASSTDCCTTCLDGVKVSASACTKCDASCETCNCA-----ATTCKAC 621
QY 862 -TFYFLWESAE-ACPLCTEHDHPIETEGACKRGFQETLYVWNEPKWCIKGISLPEKK-LAT 918
DB 622 ATGYKYTASGEGACTSC-ESDSNGVTGI-----KGCINCAAPPNNKGSVL 665
QY 919 C-----ETVDFWLKVAGVGAFATV-----LVALTCYFV 948
DB 666 CYLIKDSGSGTNSGSLSTGAIGATISAVIVVVGGLIGFLCWHP 707

RESULT 3
T10053
laminin alpha 5 chain - mouse (fragment)
C:Species: Mus musculus (house mouse)
C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 11-Jan-2000
C:Accession: T10053
R:Miner, J.H.; Lewis, R.M.; Sanes, J.R.
submitted to the EMBL Data Library, November 1997
A:Reference number: Z16923
A:Accession: T10053
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-3635 <MIN>
A:Cross-references: EMBL:U37501; NID:g2599231; PID:g2599232
C:Genetics:
A:Gene: Lama5
C:Superfamily: unassigned EGF-related proteins; EGF homology; laminin-type EGF-like homology
C:Keywords: basement membrane; cell binding; extracellular matrix
F:1888-1939/Domain: laminin-type EGF-like homology <LEG>
F:1942-1970/Domain: EGF homology <EGF>

Query Match 3.5%; Score 200; DB 2; Length 3635;
Best Local Similarity 18.5%; Pred. No. 0.00012;
Matches 174; Conservative 82; Mismatches 279; Indels 404; Gaps 46;

QY 6 RGPVRGRG-----WGRPAEAPRRGRPPSPAWICWALAGCOAWAGDLPSS---S 54
DB 1382 RGHVIGRDCSRCATGYWGFNCRP-----CDGARLDELGTGOCICP 1423

QY 55 SRPLPP-----COEKDYHFE-----YTECDSSG-SRMVAIPNSAVD---CSGLPDPVVRGKE 102
DB 1424 PRTVPPDCLVCPQSPGCHPLVGCBCNCSPGVQELTDPTCDMSDGCRCRPN-VAGRR 1482

QY 103 CTFSCASGEY-----LEMKNOV-----CSKCGEGYSLGS 132
DB 1483 CD-TCAPGEYGVPSRCPCDCHAGTMASVCDPLTGQCHCKENVQSGRCDCQVRGTFSLDA 1541

QY 133 G-----IKFEWDELPAFSGNIATF-----MDTVVGPDSRP-----164
DB 1542 ANPKGCTRCFCFATER--CGNSNLARHEFVDMEGWVLLSSROVVPVHREIELLHAD 1599

QY 165 -----DGCNNSSWIPRGVIESNRDCTVSLIYAVHLK-KSGYVFVEYQYVDNNIFFEFP 218
DB 1600 LRSVADTFSELYWOAPPYSLGDRVSSYGCTLHYELHSETQGRDIFIPYESRDPDV-----1654

QY 219 IONDOCO-----EMDTTDKWKL-TDNGEWGSHVMSKSGTNILYWR 260
DB 1655 LOGNQNSIAFLEAYPPPGQVHRGQOLQVVEGNFRHLETHNPVSRLEMMVLAGLEQLQIR 1714

QY 261 TTGILMGSKAVRPVLVKNITITIEGVAYT-----SCEFCKPGCTESN 300
DB 1715 A-----LFSQTSVSSVSLRRVLEVASAGRGPPASNVELCMCPANVYRGDSQCACAPGYRD 1770

QY 301 KPGSP--NCQVCPRTYSEK-----GAKETCRKDDSQFGSSSECTERPPCTTKDYFIHTP 355
DB 1771 TKGLFLGRCPVQCHGSHDRCLPLGSGICVGCQHNTGEGOCERC--RPGFVSDPSNPASP 1828

QY 356 CD-----EEGQTOIMYKWIPEKRICREDITDAIRLPPSEKKDCPPC 396
DB 1829 CVSCPCPLAVPSNNFADGCVLNRGRITQC-----LCR-----PGYAGASCERC 1870

QY 397 NPGFYNNG---SSSCHPC-----PPGTFSDG---TKECRPF-----PACTEPALGFE 437
DB 1871 APGFGNPLVLGSSCOPDCSNGDNPMIFSDCDPLTGACRCLRHHTGPHCERCAPGF- 1929

QY 438 YKWN-VLPGN-MKTSFCNVNGSKDGMNGWEVAGDHIQSGAGGSDNDYLIILNLHIPGFK 495
DB 1930 --YGNALLPGNCTRCDCSPCGTETCDPSG-----1957

QY 496 PPTSMGTGSELGRITTFVFETLCSADCVLFPWVDINRKSTNVESWGTREKQAYTHII 555
DB 1994 GS-----IGSDCQD-----VSGQC-NCKOGVFGKQCD 1019

DB 1958 -----RCLKAGVT-----GQCLRCLEGYFGFEQCO-----1984
QY 556 FKNATFTTFAFORTNOGQDNRFRINDMKIYSITATNAVDGVASSCRACALGSESGSS 615
DB 1985 -----GCRPCACGPAAGKSE 1999
QY 616 CVP-----CPPGHY--IEKETNOCK-----ECP-----637
DB 2000 CHPOSQCHCQPGTTGPOCLECAPGYWLPKCRQCQCPRGHCDPHTGHTCTCPGLSGE 2059
QY 638 --DTYLSIHQVYKKEACIPCGPGSKNNQ-----DHSVCYSDCFFYHEKNOILHYDFSNL 691
DB 2060 RCDTCSQQHQV-----PVPKPGGHIHCEVCHCV-----VLLD--DLE 2098
QY 692 SVGSL-----MNGPSFTSKGTKYFFHFNISLCGHEGK 723
DB 2099 RAGALLPAIREQLQGINASSAAWARLHRLNASTADLOSK 2137

RESULT 4
T23064
hypothetical protein T22A3.8 - Caenorhabditis elegans (fragment)
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000
C:Accession: T23064; T25096
R:Barlow, K.
submitted to the EMBL Data Library, October 1997
A:Reference number: Z19669
A:Accession: T23064
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2823 <WIL>
A:Cross-references: EMBL:AL008585; PIDN:CAA15432.1; GSPDB:GN00019; CESP:T22A3.8
A:Experimental source: clone H10E24
R:McMurray, A.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19980
A:Accession: T25096
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2823 <W12>
A:Cross-references: EMBL:Z81125; PIDN:CAB03385.1; GSPDB:GN00019; CESP:T22A3.8
A:Experimental source: clone T22A3
C:Genetics:
A:Gene: CESP:T22A3.8
A:Map position: 1
A:Introns: 45/1; 282/2; 312/3; 416/2; 1255/3; 1329/3; 1418/3; 1776/2; 1988/2; 2760/2
C:Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like

Query Match 3.5%; Score 199.5; DB 2; Length 2823;
Best Local Similarity 19.1%; Pred. No. 9.3e-05;
Matches 225; Conservative 110; Mismatches 363; Indels 483; Gaps 65;

QY 39 LAGCOAAWAGDLPSSSRPLPPC-----QEKDY-----HFEYTECDSSGS 78
DB 819 LGACEQC-----ECPSLDLNPNPECISTELAVLGSVSNEDNYVCINCLPLGVEGNKCEY---872
QY 79 RWRVAIPNSAVDCSG--LPDPVRGK--ECTF-----SCASGEYLEMKNQVSKC-123
DB 873 -----CSDGFEDPLTGKICETCNGNIDPMGIGNCDS-----ETGKCLKCI 914
QY 124 GEGTYSLSGSGIKFDEWDELPAFSGNIATFMDTVVGPDSRDPGCSNNSSWIPRGNIYESNR 183
DB 915 GHTTGDCSECKEHW-----CNAQLHT-----CKPCGCHTQGAVNPOCSEEN 959
QY 184 DDCVTSVLIYAVHLKSGYVFVEYQYVDNNIFFEFTIQNDQCOEMDTTTPKVKLTONGEW 243
DB 960 CEKENYIGA-----QCDRCRKHGHDVENGCPACDCNDT 993
QY 244 GSHSVMLKSGTNILYWRITGILMGSKAVKPVLVKNITIEGVAYTSECPCPKPGTFSN---300
DB 994 GS-----IGSDCQD-----VSGQC-NCKOGVFGKQCD 1019

Db 1358 TASYNGFLRFKLNEDNRRLHGRDQYFRHFPQVIFGNRIELEHHPMEINDDGIY 1417
QY 739 EI-----VAGSDDYT-----NLVGAFCQSTIIPSESKGFRAALSSQSIL 779
Db 1418 KIRLHESEMRVRSPELTUTRKMMVALQDTQGIIRGTYYTPARGDAINIEVSLDVAV 1477
QY 780 ADTFGTVVETLKNINIKEDM-----FPVPTSQDPDVHFFYKSS-----819
Db 1478 PESKIVAGLSTT-KAIGVEKLCGCPQGTGLSCQNPVEGVYRKHKREYLNQADDAIIGW 1536
QY 820 TATTSCINGRSTAVKMRNPTKSGAGVISVPSKCPAGT-----CDGCTFYFLWESAE---871
Db 1537 SEPCSC-HGHS-----QTCNPD-----TSVCTDCEHNTFGDCEHCLPGYIGDAREGGAN 1585
QY 872 -----ACPL-----CTEHDFFEIEGACKRGFOETLYVMNEPKWCIKG-ISLPEKK 915
Db 1586 ACTKACAPLVENSFSDSCVAVDHGRGYVCSCKPG-----YTGQYCETCVAGYIGDPQHI 1640
QY 916 LATCETVDFWLKVGAGVGAFTAVL-----LVALTCYFWKKNOKLEYKYSKLVMTTNS 967
Db 1641 GGTCSPCDCH-PDGLSHGACNPLSGOCECKPGVGTGRTCSMCQE-----RHAFINRVTCTSD 1695
QY 968 KECELPAAADSCAIMEGEDNEEVEVYSNKSQLLGLKLSLATK 1008
Db 1696 QCYLP-----LMETMDTMEB--HLGRQNFSG-LKPIPWK 1727

RESULT 6
T43291
laminin alpha chain - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 04-Mar-2000
C:Accession: T43291
R:Zhu, X.; Kao, G.; Joh, K.; Hall, D.H.; Wadsworth, V.K.; Hutter, H.; Vogel, B.E.; Huang
submitted to the EMBL Data Library, June 1998
A:Description: Expression, function and evolution of laminin alpha chains.
A:Reference number: 222397
A:Accession: T43291
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-3102 <ZHU>
C:Cross-references: EMBL:AF074902; PIDN:AAC26793.1
C:Genetics:
A:Map position: 1
A:Note: lama1/2
C:Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like h

Query Match 3.5%; Score 199.5; DB 2; Length 3102;
Best Local Similarity 19.1%; Pred. No. 0.0001;
Matches 225; Conservative 110; Mismatches 363; Indels 483; Gaps 65;

QY 39 LAGCQAAGDLPSSSRPLPPC-----QEKDY-----HFEYTECDSSGS 78
Db 819 LGACEQC---ECPSLDLNPNPECISTELAVLGSVASNEDNYVINCPLGYEGNKCEY---872
QY 79 RWRVAIPNSAVDCSG--LPDPVRGK--ECTF-----SCASGEYLEMKNVCSKC- 123
Db 873 -----CSDGFEDPLTGKICIECTCNGNIDPIMGIGNCDS-----ETGKCLKCI 914
QY 124 GEGTYSLGSGIKFDWDDELPAFSGFIATFMDTVVGPDSRDPDGCNNSSWIPRGVIESNR 183
Db 915 GHTTGDCSECKEHHW-----GNAQLHT-----CKPGCGHTQGANVPCSEENGE 959
QY 184 DDCVTSLIYAVHLKSGYVFFEQYVDNNIFFEPIQNDQCOEMDTTDDKVKYLTDCNEW 243
Db 960 CECKENYIGA-----QCDCKENHGDVENGCPACDCNDT 993
QY 244 GSHSVMLASGYNILYWRITGILMSKAVKPVLVKNITIEGVAYTSECPCKPGTFSN---300
Db 994 GS-----IGSDCDQ-----VSGQC-NCKQGVFGKQCD 1019
QY 301 --KPGSFN-----COVCPRNTYSEKGAECIRCKDDSDQSGSSECTERPPCT-----TKD 348

Db 1020 QCRPSYFNFTDAGQCQFCHNIY---GSIEDGKC---DQTTGKCECENVEGTMCCKCADG 1073
QY 349 YFOIHT-----PCDESGKTOIMYKWIPEK-ICREDLTDAILRPPSGEKKKDCPCPCNPGCF 400
Db 1074 YNITISGGCEDCGCDPTGSEDSVSNLVTGQCVCQKPGVT-----GLK--CDSCLPNF 1123
QY 401 Y---NNGSSSCHPCP-PGTFSDGTRKCRPCPAGTEPALGFEYKWNVLPGNNKTSFCNFVG 456
Db 1124 YGLTSEGTECEPCPAPGQVCPIDGSCVCPNT-----VGEHCNCTT---1167
QY 457 NSKCDGMNGWEVAGDHIOGAGGSDNDYILNLHLPGRKPPPTSMGTGATGSELGRITTFVFE 516
Db 1168 -----NAW-----DYHPLN-----GCK-----1179
QY 517 TLCSADCPLYEMVDINRKNSTNVNWSMGTKKEQAY-----THLPKNATFTTWAQF 568
Db 1180 -LC--DC-----SDLGSDGGMCNTFTGCKCKAAVYVGLKCDLCTHGFFNPT-----1223
QY 569 RTNOGQDNRFRINDVMKIYSITATNAVDCVASSCRACALGSGSGSSVCVPCPPGHVIEKE 628
Db 1224 -----CEPC--GCNAAGTDPDLPQCKDGOCL---1245
QY 629 TNOCKECPDPTVLSIHQVYKGEACIPCGPG-----SKNQDHSVCYSCDF-----673
Db 1246 CNEIGECP-----CKKNVHGTK-CDQCGEGTFLSDSSNLKGCTECF--CFNRTSNCEQSD 1297
QY 674 -----FYHEKEN-QILHYDPSNLSS-----VGSILMNGPS 701
Db 1298 LWQOQMYAEDRAVQEPWEFYTKKHNLNLRKESHSNYSYPTDATPLWPLPSTMLGDR 1357
QY 702 FTS-----KGTKYF-HFFNISLCGHEGKKMALCTNNITDFTVK 738
Db 1358 TASYNGFLRFKLNEDNRRLHGRDQYFRHFPQVIFGNRIELEHHPMEINDDGIY 1417
QY 739 EI-----VAGSDDYT-----NLVGAFCQSTIIPSESKGFRAALSSQSIL 779
Db 1418 KIRLHESEMRVRSPELTUTRKMMVALQDTQGIIRGTYYTPARGDAINIEVSLDVAV 1477
QY 780 ADTFGTVVETLKNINIKEDM-----FPVPTSQDPDVHFFYKSS-----819
Db 1478 PESKIVAGLSTT-KAIGVEKLCGCPQGTGLSCQNPVEGVYRKHKREYLNQADDAIIGW 1536
QY 820 TATTSCINGRSTAVKMRNPTKSGAGVISVPSKCPAGT-----CDGCTFYFLWESAE---871
Db 1537 SEPCSC-HGHS-----QTCNPD-----TSVCTDCEHNTFGDCEHCLPGYIGDAREGGAN 1585
QY 872 -----ACPL-----CTEHDFFEIEGACKRGFOETLYVMNEPKWCIKG-ISLPEKK 915
Db 1586 ACTKACAPLVENSFSDSCVAVDHGRGYVCSCKPG-----YTGQYCETCVAGYIGDPQHI 1640
QY 916 LATCETVDFWLKVGAGVGAFTAVL-----LVALTCYFWKKNOKLEYKYSKLVMTTNS 967
Db 1641 GGTCSPCDCH-PDGLSHGACNPLSGOCECKPGVGTGRTCSMCQE-----RHAFINRVTCTSD 1695
QY 968 KECELPAAADSCAIMEGEDNEEVEVYSNKSQLLGLKLSLATK 1008
Db 1696 QCYLP-----LMETMDTMEB--HLGRQNFSG-LKPIPWK 1727

RESULT 7
C42125
trophozoite cysteine-rich surface antigen 72 - Giardia lamblia (fragment)
N:Alternate names: Crp72
C:Species: Giardia lamblia
C:Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 05-Dec-1998
C:Accession: C42125
R:Adam, R.D.; Yang, Y.M.; Nash, T.E.
Mol. Cell. Biol. 12, 1194-1201, 1992
A:Title: The cysteine-rich protein gene family of Giardia lamblia: loss of the CRP170
A:Reference number: A42125; MUID:92186850; PMID:1545800
A:Accession: C42125
A:Molecule type: DNA
A:Residues: 1-677 <ADA>

Qy	20	APRRGRSPWPSPAWICWALAGCOAAWAGDLPPSSSRPLPFCOEKDYHF-----EYT	71
Dd	693	APYLKKNPSDPTGTCVSAV--DCQS-AGYTTDDSVDAKECKNACPTACAGTADKCT	750
Qy	72	ECDSGGSRWRVAIPN-----SAYDCSG-----LPDPVRGKECTFSCASGEYL	113
Dd	751	KCDANGAAPYLKKTNPSPDTGTCTCSAVDCOGSAGYITDDSVDAKECKNACPTCAGT	810
Qy	114	EMKNQCVCSC---GEGTY-----SLGSGLTKFEWD-ELPAGFSNIATFMVTGVGPSDR	163

Db 1251 VIGIIIVIMISFRNHVQDTNKSSTLADL-----SPTANNILYGTTPPVE----- 1299

Qy 611 QSGSSCVPCPP---GHVIEKE 628
|| | | | |
Db 1300 -----PPRAFYYVEDD 1311

RESULT 11

A48579
trophozoite surface protein TSP11 - Giardia lamblia
C:Species: Giardia lamblia
C:Date: 01-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 29-Jan-1999
C:Accession: A48579
R:El, P.L.; Khanna, K.K.; Manning, P.A.; Mayrhofer, G.
Mol. Biochem. Parasitol. 58, 247-257, 1993
A:Title: A gene encoding a 69-kilodalton major surface protein of Giardia intestinalis
A:Reference number: A48579; MUID:93241215; PMID:8479449
A:Contents: Ad-1
A:Accession: A48579
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-667 <EYL>
A:Cross-references: GB:M95814; NID:gi159106; PID:gi159107
A:Note: sequence extracted from NCBI backbone (NCBI:130056, NCBI:130058)

Query Match 3.28; Score 182; DB 2; Length 667;
Best Local Similarity 19.9%; Pred. No. 0.00023;
Matches 146; Conservative 58; Mismatches 236; Indels 292; Gaps 40;

Qy 103 CTFCASCEYLEMKNQVCKGEG-TYSLGSGIKPDEMDLPAGFSNTIATFMDTVVGPSPD 161
|| | | | |
Db 141 CSAPKAGEDTTPKANTCTKCAAGFLHTPSEGLSSCE-ETCPEGV-----EGHTATAESK 194
|| | | | |
Qy 162 SRPDCC-NNSWIPRGNYIESNRDDCTVSLIYAVHLKSGVYFFVYQVVDNIPFEFFIQ 220
|| | | | |
Db 195 KTCXSCTGSGSEAPNVKGI---GDCLKCM-----YNEASGNTL----- 229
|| | | | |
Qy 221 NDCQEMTTTDDKVKVLTNDGWSHVLKSGTILYKRTTG-----ILMGSKA 270
|| | | | |
Db 230 --TCEKCSAQKPSLDKTSNDNC-----TGNCAPFCSSSGGDCGDSGFLDGGNC 279
|| | | | |
Qy 271 VKPVLVKNITIEGVAYTSECFPKPTFNKPGSPNQCVPNTYSEKAKE-CIRCKDD 329
|| | | | |
Db 280 VK-----SDC-----KTENCKACT---NPKAAVECTECIST 308
|| | | | |
Qy 330 SDFSSSECTERPPCTTKDYFOIHTPCDBEETQIMYKWIIEPKICREDLTDAILRPPSC 389
|| | | | |
Db 309 HHLTPTSCVQV--COALGNVYAGTNADNK-----KACKE--CTVANCKTCD 352
|| | | | |
Qy 390 KDCPCPCNPGFYNGSS-----SCHPCPPGTFSDGTKECPCPA-----GTEPAL 434
|| | | | |
Db 353 QGOCOTCNDGDFYKNGDACSPCHESCKTCSAGTASD---CTECPTGKALKYNDGTGKTC 408
|| | | | |
Qy 435 G-----FEYKMNVLPGN-----MKTSCFNVG 456
|| | | | |
Db 409 GEGCTTGQSGACKTCGLTIDGASYCSECDTQNEYPNQNGICTSTTARTVATCKNS--NVA 466
|| | | | |
Qy 457 NSKCDG-----MNG--WEVAGDHIQSGAGSDNDYLILNLHIHIFPKPPTSWTGTATGSE 507
|| | | | |
Db 467 NGICSSCTNGFLRMNGGCVETTKFPKSVCEGANADADTCKAPVPGYK-----VE 516
|| | | | |
Qy 508 LGRITFVFETLCSADCVLYFMVDINRKSNNVSVESGGTKEQAYTHIIFKNATFTTWF 567
|| | | | |
Db 517 AGKL-----VMCKSGC-----DTGSDATTC-----TKCGGYTKI----- 546
|| | | | |
Qy 568 QRTNQCDNRREFINDMKIYISITATNAVGVASSCRACALGSEQSGSSCVPCPPGHY---- 624
|| | | | |
Db 547 -----EN-----SQTCTKC-----DSSCETCT---GAATTCVKATGYKTA 580
|| | | | |
Qy 625 IBEKTNQCKECPDPTVLSIHQVYKACIPCGPGSKNNQDHSVCYSDCFYIEKENQILII 684
|| | | | |
Db 581 LGEST--CTSCENDS-----NGVIGKGLNCAPPS----- 609
|| | | | |

RESULT 10

T25933
hypothetical protein W02C12.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T25933
R:Murray, J.; Wohldmann, P.
submitted to the EMBL Data Library, December 1996
A:Description: The sequence of C. elegans cosmid W02C12.
A:Reference number: Z20112
A:Accession: T25933
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1372 <MUR>
A:Cross-references: EMBL:U08015; PIDN:AAB37995.1; GSPDB:GN00022; CESP:W02C12.1
A:Experimental source: strain Bristol N2; clone W02C12
C:Genetics:
A:Gene: CESP:W02C12.1
A:Map position: 4
A:Introns: 29/1; 66/1; 774/2; 823/2; 1046/1; 1108/2; 1298/1

Query Match 3.28; Score 183; DB 2; Length 1372;
Best Local Similarity 20.8%; Pred. No. 0.00049;
Matches 154; Conservative 75; Mismatches 228; Indels 284; Gaps 47;

Qy 65 DYHFEYTECDSSGSR---WRVAIPNSAVDCGLPDPVRGKECTFSCAS-GEYLEMKNQV 119
|| | | | |
Db 678 EDVLYTATDNATNAQAQCNFIRVCKE--NCVDAADPNVGVQ---SCESWGP-QLKYKA 730
|| | | | |
Qy 120 CS-KGEGTYSLGSGIKPDEMDLPAGFSNTIATFMDTVVGPSPDPCGNCNNSWIPRGNY 178
|| | | | |
Db 731 CSVECRDGF-----EPRSPAFTVCAA-----DG-----KMKPNKSP 763
|| | | | |
Qy 179 TESNR-DCTVSLIYAVHLKSGVYFFVYQVDN-----NIFEFPIQDQCEMDTTDK 233
|| | | | |
Db 764 STMFRYPQCT-----KHVPATKVIVIRIYGSSPACTESSKEAFTQKVQ-OTIDAISK 816
|| | | | |
Qy 234 W--VKLTD-NGEMGS-----HSMVLSK-GTNTLY----- 258
|| | | | |
Db 817 WKMSLTDANGCVGTQVRVECGSKLPBEGRRRRNPESVLASSGVEIEIPVKRRMLVD 876
|| | | | |
Qy 259 -----WRTT-----GIL-----MGSKAVK-----PVLVKNTIEG 283
|| | | | |
Db 877 PSSGLETTIRDALHNEILLGVLNFEKVLNPNRDPVGLKIEKYLCOAGQVYVRDL---- 932
|| | | | |
Qy 284 VAYTSECPCKPGTNSNKPFSNQCVPNTYSEKAK-ECIRCKDDSQFSGS-----S 336
|| | | | |
Db 933 ---CVPCAPGTV-HSAATGECELCPIGEYQPLTARTECFKCA-PGOITASEGAISEG 984
|| | | | |
Qy 337 ECTER-PP-----CTTKDY-----FQIHTPCDEGKQTQIMYKWIIEPKICRE 376
|| | | | |
Db 985 ECKDNCPPGHQYDLSLTSVCVTCGYQYQPSAGAFEC-IPCG-IGKTLSEFATSECDRD 1042
|| | | | |
Qy 377 DLTDAILRPPSGEKDCPCNPGFYNN--GSSSCHPCPPGTFSDG----- 419
|| | | | |
Db 1043 ECPDGBQLSASGV--CQPCQIGTVRSRGENKKVACPPGTTEATMSTRRQCNTPKCK 1099
|| | | | |
Qy 420 -----TKECRPCPAGTEPALGEYK-----WNVLPG-NMKTSCF----- 453
|| | | | |
Db 1100 PGQFLVKETKNCQFCPRGT-----FQNEEQESTKLCAPDHTTAAPGATAESQCFSTNOC 1154
|| | | | |
Qy 454 -----NVGNSKCDGMNGWEVAGDHIQSGAGSDNDYLILNLHIPG 493
|| | | | |
Db 1155 ATGEYCNHANCIDLDPENDVPSECKCKPGYRGNGTHCTDAC-----NDF-CLNDGI-- 1207
|| | | | |
Qy 494 FKPPSTWATGATSELGRITFVETLCSAD-CVLYFWMVDINRK--STNVVESMGGTKEKA 550
|| | | | |
Db 1208 -----CKNNIGNVEICIKDHFSGRCELRFQASNNKLWIAIVIV-----GVV 1250
|| | | | |
Qy 551 YTHIIFKNATFTTWFATORTNQDNRREFINDMKIYISITATNAVGVASSCRACALGSE 610
|| | | | |

QY 685 YDPSNLSSVSGSLMGPSFTSKGTGYHFFNFISLCGHEGKMACTNNITDFTVKEIVAGS 744
 Db 610 -----SSTGSLV-----CYLMKGN-----TGSVNKSGSLTGAIAGISVAVV-- 649
 QY 745 DDYTNLVGAFVC 756
 Db 650 --VAGLVG-FLC 658

RESULT 12
 MMFFBI
 N: laminin beta-1 chain precursor - fruit fly (Drosophila melanogaster)
 C: Species: Drosophila melanogaster
 C: Date: 30-Jun-1991 #sequence revision 30-Jun-1991 #text_change 19-Jan-2001
 C: Accession: A28783; S14462; B28783
 R: Montell, D.J.; Goodman, C.S.
 Cell 53, 463-473, 1988

A: Title: Drosophila substrate adhesion molecule: sequence of laminin B1 chain reveals do
 A: Reference number: A28783; MUID:88210471; PMID:3365769
 A: Accession: A28783
 A: Molecule type: mRNA
 A: Residues: 1-1790 <MON1>
 A: Cross-references: EMBL:M19525
 R: Montell, D.J.; Goodman, C.S.
 submitted to the EMBL Data Library, June 1988
 A: Description: Drosophila substrate adhesion molecule: sequence of laminin B1 chain reve
 A: Reference number: S14462
 A: Accession: S14462
 A: Molecule type: mRNA
 A: Residues: 1-667, 'L', 669-725, 'VT', 728-947, 950-1790 <MON2>
 A: Cross-references: EMBL:M19525; NID:g157801; PIDN:AAA28663.1; PID:g157802
 C: Genetics:
 A: Gene: lamB1
 A: Map position: 2L 28D
 C: Complex: laminins are trimers of an alpha-type, a beta-type, and a gamma-type laminin
 C: Function:
 A: Description: interact with cells and with other basement membrane proteins to promote
 C: Superfamily: laminin beta-1 chain; laminin-type EGF-like homology
 C: Keywords: basement membrane; calcium binding; cell binding; coiled coil; extracellular
 F:1-26/Domain: signal sequence status predicted <SIG>
 F:27-1790/Product: laminin beta-1 chain #status predicted <MAT>
 F:27-288/Domain: VI <DOM6>
 F:289-561/Domain: V <DOM5>
 F:290-354/Domain: laminin-type EGF-like homology <LE01>
 F:357-417/Domain: laminin-type EGF-like homology <LE02>
 F:420-477/Domain: laminin-type EGF-like homology <LE03>
 F:480-528/Domain: laminin-type EGF-like homology <LE04>
 F:531-561/Domain: laminin-type EGF-like homology #status atypical <LE05>
 F:562-789/Domain: IV <DOM4>
 F:643-645/Region: cell attachment (R-G-D) motif
 F:790-1189/Domain: III <DOM3>
 F:791-836/Domain: laminin-type EGF-like homology <LE06>
 F:839-882/Domain: laminin-type EGF-like homology <LE07>
 F:885-932/Domain: laminin-type EGF-like homology <LE08>
 F:935-990/Domain: laminin-type EGF-like homology <LE09>
 F:968-972/Region: cell adhesion #status predicted
 F:993-1042/Domain: laminin-type EGF-like homology <LE10>
 F:1045-1093/Domain: laminin-type EGF-like homology <LE11>
 F:1096-1141/Domain: laminin-type EGF-like homology <LE12>
 F:1144-1188/Domain: laminin-type EGF-like homology <LE13>
 F:1190-1407/Domain: II <DOM2>
 F:1408-1434/Domain: alpha <ALP>
 F:1435-1790/Domain: I <DOM1>
 F:51-56/Disulfide bonds: #status predicted
 F:140,203,234,489,593,1053,1248,1303,1332,1343,1475,1517,1583,1646,1705/Binding sit
 F:1191,1194,11788/Disulfide bonds: interchain #status predicted

Query Match 3.2%; Score 181; DB 1; Length 1790;
 Best Local Similarity 18.3%; Pred No. 0.00092;
 Matches 257; Conservative 140; Mismatches 479; Indels 526; Gaps 68;

QY 39 LACQAAWAGDLPSSSRP-----LPPCQ-EKDYHFEYTECDSSCSRVRVAINP- 86
 Db 12 LALALLSOWD-PVDSORPQHGRDRPKYPPNKFRTKHPCESSCYPATGNLLIGREN 70
 QY 87 -SAVDCSGLPDPVR-----GKECTFSCASGEYLEMKNQVSKGCEGTYSLSGKIFD 137
 Db 71 LTASSTCGLHSRPERFICILSHLDKKC-FLCDTREETKHDPYKNHRIGQIILYKTKPGTNIP 129
 QY 138 EWDELPAFNSN-----IATFMDTVVGPDSRDPDGCNNSWIPRGVNI 181
 Db 130 TWMQSENGENATIQDLEAEHFTHLIITF-----TFRPAAMVIER 172
 QY 182 NRDDCTVSLIYAVHLKSGVVFPEYOVVNNIFFEFIQNDQOE-----MDITTDK 233
 Db 173 SPD-----FGTWHIYI-----FAY-----DCKSEFPQVPTVLEINIDV 207
 QY 234 W-----VKLTDNGE-----MGSHSVMLKSGTNILYWRITGILMSKAV 271
 Db 208 MCTSRYSNVSPSRNGEVIFRVLPPNINVTDPYAEHVQNOLKMTNLRITQMTKLHGLGNLL 267
 QY 272 KPVL-----VKNITIEG-----VAYTSECPCKPGTFE-----NKPQ- 303
 Db 268 DSRLENEKYIYGISNMVVRGSCSYGHASQCLPLDP-AFSQADNEDGMVHGRCECTHNT 326
 QY 304 -SFNCQVC-----PRNTYSEKGAKECIRCK-----DDSOFSGS- 335
 Db 327 KGNCECEDFFNDLPKPAFGKKTNACKKCECNDAVSHCHFEAVFTAGSFGVGGVCDN 386
 QY 336 -----SECTERPPCTTKDYFO-----IHTPCD-----EKGQITMYKWIETPKI-- 373
 Db 387 CLHNRGQHCECMFYRDPEDQITSERVQPCDCDQPGSSDDGICDLSNLEEGAVAG 446
 QY 374 ---CREDLTDALRPPSGEKKDCPPCNPGEYN---NGSSSCHPC-----PGTTFSDG----- 419
 Db 447 ACHKAFVTG-----RRCNQCKDGYWNLQSDNPECEPCTCNPLGLTNNSGVMR 496
 QY 420 -----TKECRPC-----PAGTE-----PALGFEYKMNVLPGNWKTSFEN 454
 Db 497 TGECKCKKYVTGKDCNQCPETHYGLSEPEGSLCNCDAAGSYDNCVLSGCCRPHM 556
 QY 455 VNSKCDGMNGW-----EVAGDHIQSGAGS-----DN 482
 Db 557 TGRSCSQPKQNVFIPLLPEVHEAEVVDCEISYGANGNCSLVAETPDGSGFTGIGFTRVPE 616
 QY 483 DYLILNL-HIPGFKPPTSM-----TGATGSEL 508
 Db 617 SELVFTVGDIPIRMPYDAVIRYOSTSRGDWENAFITLVRPDQVDPPEGCGEHAATSET 676
 QY 509 GRITF-----VFETLCSADCVLVFMVDINRK-----STNVVES----- 541
 Db 677 -RIPFSLPDRSRQVVALNEVCLAEAGKYKRYFERKRDVDSPTATILVDSLTLIPRID 735
 QY 542 ----MGTK-----EKOAYTHIIFKNATFTTFAFORTNQODN---RRFINVMKYISITA 591
 Db 736 VTPIFGSVLADIRKDKYKYNCKSLYDMYNADPKCNLDNLSLVFVHDGCMNCNP 795
 QY 592 TNAVDGVASSCRA-CALGSEQSSSCVPCPPG-----HYTEKET 629
 Db 796 TGSLSKVCESNGGYCOCKPNVVGROCDQCAPGTGFGPEGCKACDNCISGSKDKYCDLIT 855
 QY 630 NQCKECPDPTILSIHQVYKKEACIPCGPGSKNNQDHSVCYS-----DCFF 674
 Db 856 GOC-OCVPT-----YGRE-CNQCQPGYWNPEPCRVQCQGHAACTDPIQGTCTDC--- 904
 QY 675 YHEKENQILHYDFSNLSSVSLMNG-----PSFTSK-----GTYKF 710
 Db 905 -----QDSTTYGSCDCLDGYGNPLFGSEIGRCPRCPEVAVASGLAHADGCSLD 954
 QY 711 HFFNLSLC-----GHEGKMACTNNITDFTVKEIVAGSDDYTNLVGAFVQCSITPSES- 765
 Db 955 TRNNNMCHCQEGYSGSRCEICADNF-----FGNPDNGTCSKCECNVDLYDTG 1005
 QY 766 -----KGFRAALSSQSILADTFIGTVETTLK-----NINIKEDMFPVP 805

Db 1006 NCDROTGAACKLCLYOTGTGCHCLCKDGFQDALQOQCECDFLGTNNNTIAHCDRFTGQ 1065
QY 806 TSOIPDV-----HFFYKSTATTSCINGSTAVKMECNPTKSGAGVISVPSKCP 854
Db 1066 CCLPNVQGVRCQCAENHWKIASGEGSCNCDPIGALHEOCN---SYTG-----OCQCK 1118
QY 855 AG---TDCGCTFYFLWESAECPLCTEHDHFEIEGACKRGFQETL-----YVMNEPK 903
Db 1119 PFGGGRACNOCAHYWGNPNKQPCQCECDQFGAADFQCDRETNVCNCHGIGGYKCNE-- 1176
QY 904 WCIKISLPEKKLATC-ETVDFW-LKVGAGVGAFTAVLL-----VALTYEFWKK 950
Db 1177 -CARGVIGQPHPCSGCEGFNNWDLTSLALEDATTATILRAKEIKQVGATGAYTSEFSEL 1235
QY 951 NOKLEY-----KYSKLYWTTNSKECELPADSCAIMECEDEEEVYVSNKOS 997
Db 1236 DKKLQHIRNLLQNTSVSLVDIEKLDYETOSLRDQLOASHG-RLSETEQNLDIY--NSLS 1292
QY 998 LLG-KLKSATREKEDHFEVSQ 1018
Db 1293 LSGVELES-----QNHSLRVQ 1309

RESULT 13

T23433

hypothetical protein K08C7.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 31-Jan-2000

C:Accession: T23433

R:Berk, M.

submitted to the EMBL Data Library, March 1996

A:Reference number: 219740

A:Accession: T23433

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-3672 <WIL>

A:Cross-references: EMBL:270286; PIDN:CAA94293.1; GSPDB:GN00022; CESP:K08C7.3

A:Experimental source: clone K08C7

C:Genetics:

A:Gene: CESP:K08C7.3

A:Map position: 4

A:Introns: 66/1; 284/3; 563/1; 1187/3; 1248/3; 1300/1; 1460/1; 1623/3; 2361/3; 2988/3; 3

C:Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like H

Query Match 3.1%; Score 178.5; DB 2; Length 3672;

Best Local Similarity 22.6%; Pred No. 0.0033;

Matches 116; Conservative 54; Mismatches 203; Indels 141; Gaps 30;

QY 279 ITIEGVAYTSECPCKP--GTFENKPGSF--NCQVCPRNTYSEKGAKEICRCKDDSQFSG 334

Db 1413 LSCDCVAQGSSEFQCEYGGQCKKPGVIGRRCCERCAPGYN--PPECIKC----QCNA 1465

QY 335 SSECETRP-PCITKYDFQHTPCDEEGKQIMYKWIETPKI-----CREDLTDAILR--- 384

Db 1466 GQCDERTGQCFPPHVEGQT-CDRCVSNAGY---DPLIGQCKCGCHPQSGEGNLVCD 1521

QY 385 PPSGE-----KKDCPCNPGFYNNSSSCHPCP---PGT---FSDGTK----- 421

Db 1522 PESGOCLCRESMGGROCDRLAGFY--GPHCYGCSNCRAGTTEECDATNAOCKCKENV 1579

QY 422 ---ECRPCPAGTPEALGFYKMNVLPGNNKTSCFNVGNSKCDGMNGWEVAGDHIOGAG 478

Db 1580 YGRCACRAGT-----FDLSAENPL-GCVNCFCTGVTDCSRSSMYPVTIMSDMSFLT 1633

QY 479 GSDN-----DYLILNLHIPG-----FKPPTSMTCATGSELGRITFVFETL-----C 519

Db 1634 TDDNGMVDNKDDTVIYTSETPSNVYFNVPTEKDKYTTSYGLKLTFLKLTSTVPRGRKSM 1693

QY 520 SADCVLVFMVDINRKTNV-VESWGG---TKKQAVTHIIFKNATFTTFAFORTNOGOD 575

Db 1694 NAD-----ADVRLTGANMTIEYWASEQTNPEEQFT-VKCKLVPENFLTAEGKTVTREE 1746

QY 576 NRRFINDMWKI-----YSITATNAVGV--ASSCRACALGSEQSGS 614
Db 1747 LMKVLSLQNLITLKASYFDHPKPTSTLYEFLGISEPNGVDSVIKASSVEQCQCQCPAPYTG 1806
QY 615 SCVCPGPGHYIEKETNOCKECPDPTLYLSIHQVYGKE---ACIPC---GPGSKNNQDHVC 668
Db 1807 SCQLCASGY-----HRVQSGFLGACVPCPCNCHGHSATCDPDTGIC 1846
QY 669 YSDCFYHEKENOILHYDFSNLSSVGLMNGPSF 702
Db 1847 -TDC-----EHTNGDHCEFCNEGHYGNATNGSPY 1875

RESULT 14

T37316

probable laminin alpha chain - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Jan-2000

C:Accession: T37316

R:Joh, K.; Zhu, K.; Hedgecock, E.M.; Inoue, T.; Horl, K.

submitted to the EMBL Data Library, August 1998

A:Description: Laminin alpha chain gene in the nematode C. elegans.

A:Reference number: 221681

A:Accession: T37316

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-3704 <JOH>

A:Cross-references: EMBL:AB016806; PIDN:BAA32347.1

A:Experimental source: strain N2

C:Genetics:

A:Gene: epi-1

A:Map position: IV

A:Introns: 66/1; 284/3; 563/1; 1187/3; 1248/3; 1300/1; 1460/1; 1623/3; 2361/3; 2988/3

C:Superfamily: laminin alpha-1 chain; laminin G repeat homology; laminin-type EGF-like

Query Match 3.1%; Score 178.5; DB 2; Length 3704;

Best Local Similarity 22.6%; Pred. No. 0.0033;

Matches 116; Conservative 54; Mismatches 203; Indels 141; Gaps 30;

QY 279 ITIEGVAYTSECPCKP--GTFENKPGSF--NCQVCPRNTYSEKGAKEICRCKDDSQFSG 334

Db 1413 LSCDCVAQGSSEFQCEYGGQCKKPGVIGRRCCERCAPGYN--PPECIKC----QCNA 1465

QY 335 SSECETRP-PCITKYDFQHTPCDEEGKQIMYKWIETPKI-----CREDLTDAILR--- 384

Db 1466 GQCDERTGQCFPPHVEGQT-CDRCVSNAGY---DPLIGQCKCGCHPQSGEGNLVCD 1521

QY 385 PPSGE-----KKDCPCNPGFYNNSSSCHPCP---PGT---FSDGTK----- 421

Db 1522 PESGOCLCRESMGGROCDRLAGFY--GPHCYGCSNCRAGTTEECDATNAOCKCKENV 1579

QY 422 ---ECRPCPAGTPEALGFYKMNVLPGNNKTSCFNVGNSKCDGMNGWEVAGDHIOGAG 478

Db 1580 YGRCACRAGT-----FDLSAENPL-GCVNCFCTGVTDCSRSSMYPVTIMSDMSFLT 1633

QY 479 GSDN-----DYLILNLHIPG-----FKPPTSMTCATGSELGRITFVFETL-----C 519

Db 1634 TDDNGMVDNKDDTVIYTSETPSNVYFNVPTEKDKYTTSYGLKLTFLKLTSTVPRGRKSM 1693

QY 520 SADCVLVFMVDINRKTNV-VESWGG---TKKQAVTHIIFKNATFTTFAFORTNOGOD 575

Db 1694 NAD-----ADVRLTGANMTIEYWASEQTNPEEQFT-VKCKLVPENFLTAEGKTVTREE 1746

QY 576 NRRFINDMWKI-----YSITATNAVGV--ASSCRACALGSEQSGS 614

Db 1747 LMKVLSLQNLITLKASYFDHPKPTSTLYEFLGISEPNGVDSVIKASSVEQCQCQCPAPYTG 1806

QY 615 SCVCPGPGHYIEKETNOCKECPDPTLYLSIHQVYGKE---ACIPC---GPGSKNNQDHVC 668

Db 1807 SCQLCASGY-----HRVQSGFLGACVPCPCNCHGHSATCDPDTGIC 1846

QY 669 YSDCFYHEKENOILHYDFSNLSSVGLMNGPSF 702

Db 1847 -TDC-----EHNTNGDHCFECNEGHNATNGSPY 1875

RESULT 15

MMMSB2

laminin gamma-1 chain precursor - mouse
 N:Alternate names: laminin chain B2
 C:Species: Mus musculus (house mouse)
 C:Date: 28-Feb-1986 #sequence_revision 30-Jun-1991 #text_change 10-Dec-1999
 C:Accession: A28469; A27729; A28082; S05327; S02037; A02870; S13544; S14552
 R:Sasaki, M.; Yamada, Y.
 J. Biol. Chem. 262, 17111-17117, 1987
 A:Title: The laminin B2 chain has a multidomain structure homologous to the B1 chain.
 A:Reference number: A28469; MUID:88059118; PMID:3680290
 A:Accession: A28469
 A:Title: Primary structure of the mouse laminin B2 chain and comparison with laminin B1.
 A:Reference number: A27729; MUID:89000737; PMID:3167041
 A:Accession: A27729
 A:Molecule type: mRNA
 A:Residues: 1-1607 <SAS>
 A:Cross-references: EMBL:J03484; NID:gl98694; PIDN:AAA39405.1; PID:g293688
 R:Dutkin, M.E.; Bartos, B.B.; Liu, S.H.; Phillips, S.L.; Chung, A.E.
 Biochemistry 27, 5198-5204, 1988
 A:Title: Primary structure of the mouse laminin B2 chain and comparison with laminin B1.
 A:Reference number: A27729; MUID:89000737; PMID:3167041
 A:Accession: A27729
 A:Molecule type: mRNA
 A:Residues: 1-263, 'D', '265-336, 'C', 338-446, 'PS', 449-661, 'S', 663-885, 887-1155, 1157-1433, 'A'
 A:Cross-references: EMBL:J02930; NID:gl98702; PIDN:AAA39408.1; PID:g293691
 A:Note: The authors translated the codon TAT for residue 544 as Asp and GCG for residue
 J:Ogawa, K.; Burello, P.D.; Sasaki, M.; Yamada, Y.
 J. Biol. Chem. 263, 8384-8389, 1988
 A:Title: The laminin B2 chain promoter contains unique repeat sequences and is active in
 A:Reference number: A28082; MUID:88228071; PMID:2836421
 A:Accession: A28082
 A:Molecule type: DNA
 A:Residues: 1-215, 'A', 217-239 <OGA>
 A:Cross-references: EMBL:J03749; NID:gl98704; PIDN:AAA39409.1; PID:g554184
 R:Fujiwara, S.; Shinkai, H.; Deutzmann, R.; Paulsson, M.; Timpl, R.
 Biochem. J. 252, 453-461, 1988
 A:Title: Structure and distribution of N-linked oligosaccharide chains on various domain
 A:Reference number: S02678; MUID:88326259; PMID:24558101
 A:Accession: S02680
 A:Molecule type: protein
 A:Residues: 227-238 <FUG>
 R:Hartl, L.; Oberbaumer, I.; Deutzmann, R.
 Eur. J. Biochem. 173, 629-635, 1988
 A:Title: The N terminus of laminin A chain is homologous to the B chains.
 A:Reference number: S00624; MUID:88225080; PMID:3267223
 A:Accession: S05327
 A:Molecule type: protein
 A:Residues: 227-238; 387-393, 'F', 395-405; 881-912; 1022-1034 <HAR>
 R:Deutzmann, R.; Huber, J.; Schmetz, K.A.; Oberbaumer, I.; Hartl, L.
 Eur. J. Biochem. 177, 35-45, 1988
 A:Title: Structural study of long arm fragments of laminin. Evidence for repetitive C-ter
 A:Reference number: S01750; MUID:89030693; PMID:3181157
 A:Accession: S02037
 A:Molecule type: protein
 A:Residues: 1362-1377, 'X', 1379-1392, 'X', 1394-1406 <DEU>
 R:Barlow, D.P.; Green, N.M.; Kurkinen, M.; Hogan, B.L.M.
 EMBO J. 3, 2355-2362, 1984
 A:Title: Sequencing of laminin B chain cDNAs reveals C-terminal regions of coiled-coil a
 A:Reference number: A02870; MUID:85051302; PMID:6209134
 A:Accession: A02870
 A:Molecule type: mRNA
 A:Residues: 1391-1474, 'K', 1476-1575, 'N', 1577-1607 <BAR>
 R:Paulsson, M.; Deutzmann, R.; Timpl, R.; Dalzoppo, D.; Odermatt, E.; Engel, J.
 EMBO J. 4, 309-316, 1985
 A:Title: Evidence for coiled-coil alpha-helical regions in the long arm of laminin.
 A:Reference number: S13543; MUID:85257455; PMID:3848400
 A:Accession: S13544
 A:Molecule type: protein
 A:Residues: 1506-1523, 'X', 1525 <PAU>
 R:Olsen, D.; Nagayoshi, T.; Fazio, M.; Peltonen, J.; Jaakkola, S.; Sanborn, D.; Sasaki,
 Lab. Invest. 60, 772-782, 1989

A:Title: Human laminin: cloning and sequence analysis of cDNAs encoding A, B1 and B2
 A:Reference number: A34961; MUID:89280632; PMID:2733383
 A:Accession: S14552
 A:Molecule type: protein
 A:Residues: 881-912; 1022-1034; 1364-1377; 1379-1392; 1394-1409; 1506-1525; 1593-1606 <OLS>
 C:Genetics:
 A:Gene: Lamb-2
 A:Map position: 1
 A:Introns: 138/1; 239/3
 C:Complex: Laminins are trimers of an alpha-type, a beta-type, and a gamma-type lamin
 C:Function:
 A:Description: interact with cells and with other basement membrane proteins to promo
 C:Superfamily: laminin beta-1 chain; laminin-type EGF-like homology
 C:Keywords: basement membrane; calcium binding; cell binding; coiled coil; extracellu
 F:1-33/Domain: signal sequence #status predicted <StG>
 F:34-1607/Product: laminin gamma-1 chain #status predicted <MAT>
 F:34-283/Domain: VI <DOM6>
 F:284-502/Domain: V <DOM5>
 F:284-337/Domain: laminin-type EGF-like homology #status atypical <LE01>
 F:340-393/Domain: laminin-type EGF-like homology <LE02>
 F:396-440/Domain: laminin-type EGF-like homology <LE03>
 F:443-490/Domain: laminin-type EGF-like homology <LE04>
 F:493-502/Domain: laminin-type EGF-like homology #status atypical <LE05>
 F:503-687/Domain: IV <DOM4>
 F:688-1032/Domain: III <DOM3>
 F:688-719/Domain: laminin-type EGF-like homology #status atypical <LE06>
 F:722-768/Domain: laminin-type EGF-like homology <LE07>
 F:771-823/Domain: laminin-type EGF-like homology <LE08>
 F:826-879/Domain: laminin-type EGF-like homology <LE09>
 F:882-930/Domain: laminin-type EGF-like homology <LE10>
 F:933-978/Domain: laminin-type EGF-like homology <LE11>
 F:981-1026/Domain: laminin-type EGF-like homology <LE12>
 F:1033-1607/Domain: II/I <DOM2>
 F:1033-1607/Region: heptad repeats
 F:38-48/Disulfide bonds: #status predicted
 F:58,132,574,648,1020,1105,1159,1173,1203,1221,1239,1437/Binding site: carbohydrate (C
 F:1029,1032/Disulfide bonds: interchain #status predicted
 F:1378,1393/Binding site: carbohydrate (Asn) (covalent) #status experimental
 F:1598/Disulfide bonds: interchain (to chain B1) #status experimental

Query Match 3.1%; Score 177.5; DB 1; Length 1607;

Best Local Similarity 19.3%; Pred. No. 0.0014;

Matches 215; Conservative 115; Mismatches 437; Indels 347; Gaps 57;

QY 13 GWCRAEAPR-RCRPPNSPANICWALACQAAWAGDLPPSSSRLLPPCQEKDYHFET 71

Db 3 GGGRAALQPRGR-----LWPLLAVALAAVAGCVRAAMD----- 36

QY 72 ECDSSGSRWRVAIP---NSAVD-----CSGLPDPVRGKCTFSCASGEYLEMKNQVC 120

Db 37 ECADEGGRPORCMPEFVNAAFNVVATWTCGTPPE----EYCVQGVVG-----VTKSC 87

QY 121 SKCGEGTYSLGSGIKF--DEWDELPAFGSNIAFMDFMTVVGPSD----- 161

Db 88 HLCDAQQHLQHGAAFLTDYNNQADTTWQSOTLAGVQYPSNLTLLHGLKAFDITYVR 147

QY 162 -----SRPDGC-----NNSSWIP-----RGNYIESNR-----DD-----CT--- 187

Db 148 LKFTSRPESFAIYKRTREDGWPVQYVYSGCENTYSKANRGFIITGGDEQALCTDEF 207

QY 188 --VSLIYAVHL-----KKGYVFVEYVDNNTIFFEFFFIONDOCOEMDTTDDKWKVLT 238

Db 208 SDISPLTGGNVAFTLEGPSAYNF-----DN-----SPVLQEWVATDTRVTLN 252

QY 239 DNGENG----SHSVMLKSGNTLYWRTTILMG-----SKAVKPVLVKNI-TIEGV 284

Db 253 RLNTFGDEVFNEPKVLKS----YYVAISDFAVGRCCKNGHASECVKNEFDKLMCKNKH 308

QY 285 AYTSECFCKPGCTFSNKP-----GTFNQCVCPRNTYSEKAKECIR 325

Db 309 TVGVDCXCKCLP--FFNDRPMRRATAESASLPCDCNGRSQECYFDPELYRSTGHGHCTN 367

QY 326 CKDDSQFSGSSECTER-----PPCTTKDYFIQIHTPCDEEGKTKIWKWIEPKICR 375

```
Db 368 CRDNTDGAKCERCRNFRLNGNTEACSPCHCSVGLSLTQCDSDYGRCS-----CK 417
Qy 376 EDLTDAIRLPPSGEKKDQPPCPNPGYNNNGSSCHPCPGTFSDTGTECR-----PCPAG 429
Db 418 PCV-----MGDK--CDRCOPGFHSLTEACRCP-SCDLRGSTDECNVETGRVCVKDN 466
Qy 430 TEPALGFYK-----WNVLPQNMK--TSCFNVGNSK-CDGMNGWEV-----AG 470
Db 467 VE--GFNCERCKPGFFNLESSNPKGCTPCFCFHSSVCTNAVGYSVYDISSTFOIDEDG 523
Qy 471 DHIQSGAG-----GSDND--YLILNLHIPG-FKPPTSMTCATGSELGR-ITVFV-----515
Db 524 WRVEQRDSEASLEWSSRDIAVISDSYFPRIYFAPVKFLGNQVLVSYGQNLFSFSFRVDR 583
Qy 516 -ETLCSADCVLFYVDINRKSNNVNESGGTKEKQAYTHIFKNATFTTW-----AF 567
Db 584 RTRLSAEDLVLEGAGL-RVSVPLIAQNSYSETVKYIFRLHEATDYPWRPALSPREF 642
Qy 568 QRTNOGQDNRRFINDMVKI-----YSITATNAVGVASSCRACALGSEQSGSCVPCPPG 622
Db 643 Q-----KLLNLTSTIKIRGTYSERTAGYLDV-----TLQSARPG-----PGVPA 682
Qy 623 HYIEKTNCKECPDPTYLSIHQVYVGEACIPCGFGSKNNQDHVSVYSDCFY-----675
Db 683 TWVESCT-----CPVG-----YGGQFCETCLPGYRRETSLGYPSPCVLCTCNHSE 729
Qy 676 -----HEKENQILHYDFSNLSSVGLMNGPSTSKGTYKFFHFNISL 717
Db 730 TCDPETGVCDRDNAGPCEKCDGYGDSLTGTSDDCQPCPCGGSSCAIVPKTEVV 789
Qy 718 CGH-----EGKXALCTNNITDFTVKEIVAGSDDYTNLVGAFVCSOTIIPSESKGFRAAL 772
Db 790 CTHCPTGTAGKRCCLDDGYFGDPL-----GSGPVRLCRPCOCNDNIDPN-AVGNCRNL 843
Qy 773 SSQSILADTFTGVTVEITLKNINIKEDMFPVPTSQIPDVHFFYKSSATTSCINGRSTA 832
Db 844 TGECLKCIYNTAGFYCD-----RCKEGFFGNPLAPNP-----ADKCKACACNPYGTVO 891
Qy 833 VKMRCNPTKSGAGVISVPSKCPAGTCDGCTFYFLWESAEACPLCTEHDHPHEIAG--KR 890
Db 892 QOSSCNPTVGQCQCLPHVNSGRDCGTC--PGYINLQSGGCGCCDCHALGSTNGQCDIRT 949
Qy 891 GFOETLYVWNEPKWICKIGISLPEKKLATCETVDF 924
Db 950 GQCE-----CQGIT--GQHCERCETNHF 971
```

Search completed: May 12, 2003, 13:21:31
Job time : 49.9327 secs

GenCore version 5.1.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 12, 2003, 11:12:58 ; Search time 20.6432 Seconds

(without alignments)

2063.447 Million cell updates/sec

Title: US-10-073-333A-4

Perfect score: 5681

Sequence: 1 MLFRAGPVVRGNGRPAEA.....REKEDHFESVLKTSRSPNI 1027

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 segs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	248.5	4.4	1696	1	PCK5_BRACL
2	242.5	4.3	1877	1	PCK5_MOUSE
3	223.5	3.9	687	1	VS41_GIALA
4	210	3.7	713	1	TS44_GIALA
5	200	3.5	3718	1	LMA5_MOUSE
6	194	3.4	5376	1	ZAN_MOUSE
7	184	3.2	1790	1	LMB1_DROME
8	182	3.2	667	1	TS11_GIALA
9	178.5	3.1	3672	1	LM2_CAEEL
10	177.5	3.1	1607	1	LMG1_MOUSE
11	177.5	3.1	2907	1	FBN2_MOUSE
12	173.5	3.1	1581	1	LMG3_MOUSE
13	173.5	3.1	2871	1	FBN1_MOUSE
14	173	3.0	1786	1	LMB1_MOUSE
15	170.5	3.0	2437	1	NFC1_BRARE
16	170	3.0	1557	1	LM2_CAEEL
17	170	3.0	1786	1	LMB1_HUMAN
18	170	3.0	2491	1	MPRI_HUMAN
19	169	3.0	1639	1	LMG1_DROME
20	167.5	2.9	1587	1	LMG3_HUMAN
21	166	2.9	2871	1	FBN1_HUMAN
22	165	2.9	3110	1	LM2_HUMAN
23	164	2.9	2871	1	FBN1_BOVIN
24	164	2.9	2871	1	FBN1_PIG
25	164	2.9	3084	1	LM1_MOUSE
26	161	2.8	3712	1	LMA_DROME
27	160.5	2.8	1609	1	LMG1_HUMAN
28	160.5	2.8	2911	1	FBN2_HUMAN
29	158	2.8	3695	1	LMA5_HUMAN
30	157.5	2.8	1219	1	JAG1_RAT
31	157	2.8	2531	1	NFC1_MOUSE
32	156.5	2.8	3106	1	LM2_MOUSE
33	155.5	2.7	2003	1	NTC4_HUMAN

34	155	2.7	2499	1	MPRI_BOVIN	P08169 bos taurus
35	154.5	2.7	2482	1	VWF_PIG	Q28833 sus scrofa
36	153.5	2.7	1680	1	FUR2_DROME	P30432 drosophila
37	153	2.7	1959	1	AGRI_RAT	P25304 rattus norv
38	153	2.7	2318	1	NTC3_MOUSE	Q61982 mus musculus
39	152	2.7	1218	1	JAG1_MOUSE	Q94xx0 mus musculus
40	152	2.7	3707	1	PGBM_MOUSE	Q05793 mus musculus
41	151.5	2.7	2703	1	NOTC_DROME	P07207 drosophila
42	151	2.7	1231	1	CFAR_HUMAN	P08603 homo sapien
43	151	2.7	2470	1	NTC2_MOUSE	Q35316 mus musculus
44	149.5	2.6	2524	1	NOTC_XENLA	P21783 xenopus lae
45	149	2.6	2321	1	NTC3_HUMAN	Q9um47 homo sapien

ALIGNMENTS

RESULT 1
PCK5_BRACL
ID PCK5_BRACL STANDARD; PRT: 1696 AA.
AC Q9NJ15; Q9NJ16; Q9NJ14;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Proprotein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)
DE (Proprotein convertase PC6-like) (apc6).
GN PC6.
OS Branchiostoma californiensis (California lancelet) (Amphioxus).
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC Branchiostoma.
OX NCBI_TaxID=7738;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS A; B AND C).
RX MEDLINE=20175281; PubMed=10708868;
RA Oliva A.A. Jr., Chan S.J., Steiner D.F.;
RT "Evolution of the prohormone convertases: identification of a
homologue of PC6 in the protochordate amphioxus";
RL Biochim. Biophys. Acta 1477:338-348(2000).
CC -!- FUNCTION: LIKELY TO REPRESENT A WIDESPREAD ENDOPEPTIDASE ACTIVITY
WITHIN THE CONSTITUTIVE AND REGULATED SECRETORY PATHWAY. CAPABLE
OF CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR
PROTEIN BY CLEAVAGE OF ARG-XAA-YAA-ARG-1-ZAA BONDS, WHERE XAA
CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.
CC -!- SUBCELLULAR LOCATION: ISOFORM A AND ISOFORM C ARE SECRETED.
ISOFORM B IS A TYPE I MEMBRANE PROTEIN.
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; A, B (SHOWN HERE) AND C; ARE
PRODUCED BY ALTERNATIVE SPLICING.
CC -!- DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE
ASSISTING THE FOLDING OF THE ZYMOGEN WITHIN THE ENDOPLASMIC
RETICULUM.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
CC -!- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF184615; AAF26300.1;
EMBL; AF184616; AAF26301.1;
EMBL; AF184617; AAF26302.1;
HSSP; Q99405; IMPT.
MEROPS; S08.UPB;
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002884; P_domain.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 1.
DR Pfam; PF01483; P; 1.


```
FT VARSPLIC 916 1877 (IN ISOFORM PC5A).
SQ SEQUENCE 1877 AA; 209287 MW; EC850E2DF20EA1C3 CRC64;

Query Match 4.3% Score 242.5; DB 1; Length 1877;
Best Local Similarity 19.1%; Pred. No. 1.1e-09;
Matches 221; Conservative 97; Mismatches 382; Indels 456; Gaps 64;

Qy 61 CQKDYHFEYTECDSSGRWVAIPNSAVDCSGLPDPVRGKEC-----TFSCA 108
Db 676 CPPGHYHAKKRC-----RKCAPN-CESCFG-----SHGQCLSKYGYFLNEETSCV 723
Qy 109 S-----GEYLEMKNVCKSGEGTSLGSGIKFDEWDELPAFGS-----NIATEMDTVVG 158
Db 724 TQCPGYSYEDIKKNVCGKSENCAC-----IGFHNCTCKGGLSLGSRCSVTCEGQGFEN 780
Qy 159 PSDSRP-----DCGNSSWTPRGNYTESNR--DDCTVSLIYAVHLKKSGYVF 203
Db 781 GHDCQPCHRFCATCSGAGADGCTNCT--EGVMEEGRVQSCSVS--YLLDHSSEGGYK- 835
Qy 204 FEYQVVDNIEF-----PTQEDICISCPVTRVLDGRC-----VMNCPSNKFEEKKQCH 935
Db 292 PCKPOTENKPGSPNQCVRPNYSEKAKECIRKDD-----SQFSGSSECTERPPCTTK 347
Db 936 PC-----HYTCOGC-----QSGSPNCTSCRADKHGOERFLYHGELEN--CPVG 978
Qy 348 DY-FQIHT--PCDEEGKQIMYKWKPEKTCREDLTDALRLPS-----GE 389
Db 979 HYPAGKHTCLPCPD--NCELCY---NPHICRCMGGYVLIIPNHTCQKLECRQGEFQDSE 1033
Qy 390 KKDCCPPCPGFY---NGSSSCHPCPPGTFSDGTKECRPCPAGTAPALGFYKKNVLP 446
Db 1034 YEECPMBEGCLGCTEDDPGATCATGYMPERHCYKACPEKT-----FGVKECRACG 1088
Qy 447 NTKTSFVNGSKCDGMNCGWEVAGDHIQSGAGSDNDYLLNLHLPGRKPPPTMTGATGS 506
Db 1089 ---TNGSCDQHEC---YWCBEFGFLSGSGCVQDCG-----PGH-----GDO 1125
Qy 507 ELGRITVFETLCSADCVLYFMVDINRKST--NVVESMGKTWEKQAYTHIIFKNATFTFT 564
Db 1126 ELGEC-----KPCRACETCTGSGYNQCSQCGELQHLHGHTCLWSTWPQVEGKD----- 1174
Qy 565 WAFQRTNQG-----QDNRRFNDVMKYISITATNAVGVASSCRACALGSEQ 611
Db 1175 W-----NEAVPTKEPSLVRSLQLDRKW---KVQI-KRDATSONQPCFHSCKTC----- 1219
Qy 612 SGSSCVPCPPGHI-----EKTNOCKEC-----PPD 638
Db 1220 NGLSCASCTGMWLYLQACVPSPQGTWPSVTSGCEKCEDCVSCGADLCCQCLSQPD 1279
Qy 639 TYLSIHO-----VYKGE-ACIPCGPGSKNNO-----DHSVCY----- 669
Db 1280 NTLLLEGRYHSCPGFPAKDVGCVEHCSPCKTCBGNATSCNSCGDFVLHDGVCKWTC 1339
Qy 670 -----SDCFYHEK-----ENOILHYDFSNSVSGSL----- 696
Db 1340 PEKHVAVEGVCKHCPERCQDCI--HEKTKCECMPPDFFLYNDMCHRSCKPSFYPMRQCVP 1397
Qy 697 -----MNCPSFTSGTYFHFNFSLCCHGCKK--ALCTNNITDFTVKE----- 739
Db 1398 CHKNLECGPKED-----DCKVCADTSKALHNLGLDCECPGYKKEENDECRD 1447
Qy 740 -----IVAGSDDYTNLV-----GAFVCQSTIIPSESIGFR 769
Db 1448 CPESCLICSSAWTCLACREGFTVVHDVCTAPKECAAVEYWDGSHRCQ---PCHKKCSR 1503
Qy 770 AALSSQSIILADTFIGVTVETTLKNIN--IKEDMFPVPTSOIPDVHFFYKSSATTATSC--- 825
```

RESULT 3

```
VS41_GIALA STANDARD: PRT: 687 AA.
AC P92127;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Variant-specific surface protein VSP4AI precursor (CRISP-90).
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardinales; Giardia.
OX NCBI_TaxID=5741;
RN 1;
RP SEQUENCE FROM N.A.
RC STRAIN=O2-4AI;
RX Papanastasiou P., Bruderer T., Li Y., Bommell C., Koehler P.;
RA "Primary structure and biochemical properties of a variant-specific
RT surface protein of Giardia."
RL Mol. Biochem. Parasitol. 86:13-27(1997).
RN 12;
RP CHARACTERIZATION.
RX MEDLINE=97233006; PubMed=9078242;
RA Papanastasiou P., McConville M.J., Ralton J., Koehler P.;
RT "The variant-specific surface protein of Giardia, VSP4AI, is a
RL glycosylated and palmitoylated protein."
Biochem. J. 322:49-56(1997).
CC 1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ANCHORED TO THE
PLASMA MEMBRANE.
CC 1- PTM: O-GLYCOSYLATED. THE MAJOR GLYCAN IS A TRISACCHARIDE WITH GLC
AT THE REDUCING TERMINUS.
CC 1- PTM: PALMITOYLATED.
CC 1- SIMILARITY: BELONGS TO THE GIARDIA VARIANT SURFACE PROTEIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL; Z83743; CAB06038.1;
CC HSP; O14763; 1D0G.
DR GlycoSuiteDB; P92127;
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR005127; Giardia_VSP.
DR Pfam; PF03302; VSP; 2.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00001; EGF_Like; 1.
DR SMART; SM00261; FU; 3.
KW Antigen; Repeat; Transmembrane; Glycoprotein; Lipoprotein; Palmitate;
Signal.
FT SIGNAL 1 14
FT CHAIN 15 687
FT DOMAIN 15 660
FT TRANSMEM 661 681
FT POTENTIAL.
FT VARIANT-SPECIFIC SURFACE PROTEIN VSP4AI.
FT EXTRACELLULAR (POTENTIAL).
FT POTENTIAL.
```

```
FT DOMAIN 682 687 CYTOPLASMIC (POTENTIAL).
SQ SEQUENCE 687 AA; 70857 MW; D892F675D626D7EC CRC64;

Query Match
Best Local Similarity 3.9%; Score 223.5; DB 1; Length 687;
Matches 144; Conservative 88; Mismatches 255; Indels 283; Gaps 36;

QY 290 CFCPCPGTF-----SNKPSGNCQVCPRTYSEKGAKEICRCKDDSQSGSSECTE 340
D 83 CGKCGDGVFLFMGCGYKTESQGS---ETC-----TTASNGL--CTACKVDSQY----- 126
QY 341 RPPCTTKDYFQIHTCDDEGKTQIMYKIEPKICREDLTDALRLPSPSEKKDCPCPNPGF 400
D 127 -----IFQNKATPSEKSGEILC-W-----DITD--RNGVMG-VANCATCTAPA 166
QY 401 YNNGSSCHPCPPGTFSDGTRKCRPCPAGTALPGEYKWNVLGNMKTSCFNVGNSKC 460
D 167 SSTGPTATCEMAGTYKSDTECAACHDC-----ATCSGANNQC 207
QY 461 DGMNGEIVAGHIQSGAGSDNDYLILNLHPGKPTSMT--GATGSELGRITFVFET- 517
D 208 TSCE-----TGKYLKSNQCEKN-----TCNTNHYPDTSMTCTVACTVLDANCACTSDSA 258
QY 518 -----LCSADCVLYFWDINRKNSTNVVSGGTEKQAYTHIIPKNTATFTTWFORT 570
D 259 TAKGKCLTCSNKPRTILD--GTSTCVENSYAGC----- 291
QY 571 NOGQDNRFPIND--WVKIYSITATNAVGVASSRACALGSEQSGSCVCPGPGHYIEK 627
D 292 -QGADNELFMKEDQSACLLCGDTKEASNDKGVANCRCTCKNANDSPPTCTACLDGFLER 350
QY 628 ET-----NCKCEPDDTYLSIHQVYKEACIPC----- 655
D 351 GSCITTCADNCACTSEATTEOKICKAGFFLASP---GEGKICISCDTNNGGIDGCAEC 407
QY 656 -----GP-----GSKNN-----QDHSVCYSDCFYHEKENOILHYDSN 689
D 408 TKEPAGPLKTKCKNRPAGTSDNYTCTEKTCDPTVC----- 446
QY 690 LSSVGLMNGPFSFTSGTKYKTHFFNISLCHEGKKM--ALCTNN-----ITDFTVKEIV 741
D 447 -GGTSGACDAIVIDANGREHY--CSYCGETNKFPIDGLCTDNGKTAGCTDHTCSYCA 502
QY 742 AGSDDY-----TNLGAFCVQSTI-----IPSEKGF-----RAALSSOSIILADTFI 784
D 503 AGFFLYMGCGYKIDTVPGSYMCSKSTTAGVCDTPNANNRFFVVPKRAISAEQSVLACGNPL 562
QY 785 GVTVETTLKNIKEDMFPVPTSQIPDVHFFYKSSAT-TSCINGRSTAVKRCNPTKSG 843
D 563 G-----TIAGGNAYVGVGCGSQCTAPDARAGGMATVCTACEDGK-----KPGKSG 609
QY 844 AGVISVPKCPAGTCGCTFYFLWESAEACPLCTEHDFHEIEGACRKGFOETLYVWNEPK 903
D 610 TQCVV---CPDANKSCT-----MDDVCEE----- 631
QY 904 WKIKGISLPEKKLATCEVDFWLKVGAGVGAFTAVL-----LVALTCYFM 948
D 632 -CADGFSLDNGKCVSSGNTKSGLSGTGATAGISVAIVVVGGLVGLFLCWWF 680

RESULT 4
TS4A_GIALA
ID TS4A_GIALA STANDARD; PRT; 713 AA.
AC P21849;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Major surface labeled trophozoite antigen 417 precursor.
GN TSA 417.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardia.
OX NCBI_TaxID=5741;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 30957 / NB;
RA MEDLINE=90280395; PubMed=2352929;
RA Gillin F.D., Hagblom P., Harwood J., Aley S.B., Reiner D.S.,
RA McCaffery M., SO M., Guiney D.G.;
RT "Isolation and expression of the gene for a major surface protein of
RT Giardia lamblia.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:4463-4467(1990).
RN [2]
RP SEQUENCE OF 480-620 FROM N.A.
RC STRAIN=AD-1;
RX MEDLINE=93314970; PubMed=8325510;
RA Ey P.L., Mayrhofer G.;
RT "Two genes encoding homologous 70-kDa surface proteins are present
RT within individual trophozoites of the binucleate protozoan parasite
RT Giardia intestinalis.";
RL Gene 129:257-262(1993).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ANCHORED TO THE
CC PLASMA MEMBRANE.
CC -!- DOMAIN: CONTAINS 29 REPEATS OF THE CXXC MOTIF.
CC -!- SIMILARITY: BELONGS TO THE GIARDIA VARIANT SURFACE PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M33641; AAA02688.1; -
CC EMBL; M97488; AAA02581.1; -
CC PIR; A35502; A35502;
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR005127; Giardia_vsp.
CC Pfam; PF03302; VSP; 2.
CC SMART; SM00181; EGF; 1.
CC SMART; SM00001; EGF-like; 1.
CC SMART; SM00261; FU; 3.
CC Signal; Antigen; Glycoprotein; Transmembrane; Repeat.
KW SIGNAL
FT CHAIN 1 17
FT CHAIN 18 713
FT FT MAJOR SURFACE-LABELED TROPHOZOITE
FT ANTIGEN 417.
FT DOMAIN 18 679
FT EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 680 708
FT DOMAIN 709 713
FT CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 289 289
FT CARBOHYD 676 676
FT N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 582 582
FT A -> T (IN STRAIN ADELAIDE-1).
FT VARIANT 606 606
FT A -> S (IN STRAIN ADELAIDE-1).
SQ SEQUENCE 713 AA; 72510 MW; 9AD7195843DE5601 CRC64;

Query Match 3.7%; Score 210; DB 1; Length 713;
Best Local Similarity 19.7%; Pred. No. 7.9e-08;
Matches 186; Conservative 96; Mismatches 300; Indels 360; Gaps 53;

QY 90 DCSGLP-DPVRGKECTFSCASGEYLEMKNQ-----VCSKCG-----ECTYSLGSGIK--- 135
D 43 ECNGANYAPVNG-QCDVNAEGPSKTLCPQHSAGKCTCGGNSFMKDGCVSSGEGELPGH 101
QY 136 ----FDEWDEL-----PAGFSN-----IATFMDTVVGPSPD---S 162
D 102 SILCLSSDGDGVCTEAPAGYFAPVGAANTEQSVIACGDTGTGTIAGGNTYKGIADCAECS 161
QY 163 RPDGNNSSWTPRGNYIESNRDDCTVSLIYAVHLKKSGYVFVEYQYVDNNIFFEFFFQND 222
D 162 APDA---TAGAAGKAVATCTK--CGVS-----KYLKDNVCV-----D 193
QY 223 OCQEMDTTDTKWKVLTNDNGENGSHVMLKSGTNILYWRPTTGLMGSKAVKPVLVKNITIE 282
D 194 KAQCNSGNTNKFVAV-DDSENGNKCVCSDNLN-----G 226
```

```
Oy 283 GVAYTSECFCKPGTGNKPSFNCQVCPNRYSEKGAKECIRCKDDSOFGSSE---C 338
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 227 GVA-----NCDTCT---SYDEQSKKIKCTCTDNNLYLKTTSSEGTSC 263
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 339 TERPCTTKDYFOIHTPCDEBEKTKQIMYKWIPIKICREDLTDALRLPSPGKCKDCPPCNP 398
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 264 VKQDQC---KDGFE---FPKDDSS-----AGNK---CLPCND 290
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 399 GFYNNSSSCHPCP---PCTESDGTKECRPCPAGTEPALGFEBYKWNVLPGNMKTSFCFNVG 456
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 291 S---TDGIANCATCALVSRSCAALVTCACDYGKPSAD-----KTCCEAVS 335
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 457 NSKCDMGWVAGDHIOQAGGSDNDYLILNLHHPGKFPPTSMGTGATGSELGRITTFVFE 516
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 336 NCKTPGCRACSEKNEVEVCTDCDGSSTVL-----TPTSQ----- 369
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 517 TICSADCVLYFWNDLNRKSTNVVSWGGTKEKQAYTHIFKNATFTFWAFORT---NOGO 574
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 370 --CIDSCA-----KIGNYVGATEGAKK-----LCRECTA-----ANCKTCDGQ 407
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 575 DNRREFINDWKIYSITATNADVGVASSCRACALGSEQSGSCVPCPPGHYIEK-----ET 629
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 408 -----COACNDGFYKNGDACSPC---HESCKTCSACTA 437
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 630 NCKCEPPTDYLISIHQVYKKEACI--PCGPGSKNNDHVSVCYSDCFYHEKENQILHYDFS 688
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 438 SDCTECPGKALR-----YCDGCTKCTGEGCTTGTGAGAC----- 473
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 689 NLSVSGSLMNGPSTSKCTKYFHFENISLGHGKMKALCTNNITDFVKEIVAG---SDD 746
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 474 --KTGGLTIDGASYCEGATTEYPONGVC---APKARATPTCNDSPIONGVCTCADN 528
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 747 YTNLVGAFVCSQTIIPSESKGFRALSSQSIILADTFIGVITVETTLKLNINKEDEMPVPT 806
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 529 YPKMNGG--CYETV---KYPGKTVCSISAPN-----GGTCQKAADGYKLDGSLTVCS 575
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 807 SOIPDVHFFYKSTATTSCING--RSTAVKMRCN---PTKSGAGVISVPSKCPAGTCDGC 861
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 576 EGCKEC-----ASSTDCCTCLDGYKVSACTKCDASCETCNGA-----ATTCKAC 621
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 862 -TFYFLWESAE--ACPLCTEHDFHEGACKRGFTLVVWNEPKWIKIGISLPEKK--LAT 918
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 622 ATGYKYTASGEACATSC--ESDSNGVTGI-----KCLNCAPPNNKGSVL 665
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 919 C-----ETVDFWLKVGAGVGAFTAVL-----LVALTCYFW 948
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 666 CVLKIDSGSTNKSGLSTGAIGASIVAVIVVGGLLIGFLCWNF 707
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
LMA5_MOUSE STANDARD: PRT: 3718 AA.
AC Q61001; Q9JH06;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Laminin alpha-5 chain precursor.
GN LMA5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_taxid=10090;
RX [1]
RP SEQUENCE OF 1-92 FROM N.A.
RA Timpi R., Sasaki T.;
RT "Completion of the N-terminal sequence of the murine Laminin alpha 5
RT chain.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 84-3718 FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Lung;
RX MEDLINE=96081906; PubMed=7499364;
RA Miner J.H., Lewis R.M., Sanes J.R.;
```

```
RT "Molecular cloning of a novel laminin chain, alpha 5, and widespread
RT expression in adult mouse tissues.";
RL J. Biol. Chem. 270:28523-28526(1995).
RN [3]
RP REVISIONS.
RA Miner J.H., Lewis R.M., Sanes J.R.;
RA Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RL -1- FUNCTION: Binding to cells via a high affinity receptor, laminin
CC is thought to mediate the attachment, migration, and organization
CC of cells into tissues during embryonic development by interacting
CC with other extracellular matrix components.
CC -1- FUNCTION: ALPHA-5 CHAIN MAY BE THE MAJOR LAMININ ALPHA CHAIN OF
CC ADULT EPITHELIAL AND/OR ENDOTHELIAL BASAL LAMINAE.
CC -1- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
CC different polypeptide chains (alpha, beta, gamma), which are bound
CC to each other by disulfide bonds into a cross-shaped molecule
CC comprising one long and three short arms with globules at each
CC end.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR; FOUND IN THE BASEMENT
CC MEMBRANES (MAJOR COMPONENT).
CC -1- TISSUE SPECIFICITY: IN ADULT, HIGH LEVELS IN HEART, LUNG, AND
CC KIDNEY; LOWER IN BRAIN, MUSCLE AND TESTIS; VERY LOW IN LIVER, GUT
CC AND SKIN. EXPRESSED IN MANY TISSUES IN EMBRYONIC DAY 11.
CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC -1- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.
CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -1- SIMILARITY: CONTAINS 21.5 LAMININ EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.
CC -1- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ293593; CAB99255.1; -.
DR EMBL; U37501; AAC53430.1; -.
DR HSSP; P02468; 1TLE.
DR MGD; MGI:105382; Lama5.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001886; LamNT.
DR InterPro; IPR000034; Laminin_B.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR001791; Laminin_G.
DR Pfam; PF00052; laminin_B; 1.
DR Pfam; PF00053; laminin_EGF; 19.
DR Pfam; PF00054; laminin_G; 2.
DR Pfam; PF00055; laminin_Nterm; 1.
DR ProDom; PD002082; LamNT; 1.
DR ProDom; PD003031; Laminin_B; 1.
DR PROSITE; PS00022; EGF_1; 19.
DR PROSITE; PS01186; EGF_2; 3.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 19.
DR PROSITE; PS00025; LAM_G_DOMAIN; 5.
DR Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
FT SIGNAL 1 40
FT CHAIN 41 3718 LAMININ ALPHA-5 CHAIN
FT DOMAIN 41 304 LAMININ N-TERMINAL (DOMAIN VI).
FT DOMAIN 305 363 LAMININ EGF-LIKE 1.
FT DOMAIN 364 433 LAMININ EGF-LIKE 2.
FT DOMAIN 434 479 LAMININ EGF-LIKE 3.
FT DOMAIN 500 546 LAMININ EGF-LIKE 4.
FT DOMAIN 547 592 LAMININ EGF-LIKE 5.
FT DOMAIN 593 637 LAMININ EGF-LIKE 6.
FT DOMAIN 638 682 LAMININ EGF-LIKE 7.
FT DOMAIN 683 728 LAMININ EGF-LIKE 8.
FT DOMAIN 729 781 LAMININ EGF-LIKE 9.
FT DOMAIN 782 833 LAMININ EGF-LIKE 10.
```

FT DOMAIN 834 855 LAMININ EGF-LIKE 11 (INCOMPLETE).
FT DOMAIN 856 1442 LAMININ DOMAIN IV 1 (DOMAIN IV B).
FT DOMAIN 1443 1488 LAMININ EGF-LIKE 12.
FT DOMAIN 1489 1532 LAMININ EGF-LIKE 13.
FT DOMAIN 1533 1581 LAMININ EGF-LIKE 14.
FT DOMAIN 1582 1632 LAMININ EGF-LIKE 15.
FT DOMAIN 1633 1842 LAMININ EGF-LIKE 16 (N-TERMINAL).
FT DOMAIN 1843 1831 LAMININ DOMAIN IV 2 (DOMAIN IV A).
FT DOMAIN 1832 1864 LAMININ EGF-LIKE 16 (C-TERMINAL).
FT DOMAIN 1865 1914 LAMININ EGF-LIKE 17.
FT DOMAIN 1915 1970 LAMININ EGF-LIKE 18.
FT DOMAIN 1971 2024 LAMININ EGF-LIKE 19.
FT DOMAIN 2025 2071 LAMININ EGF-LIKE 20.
FT DOMAIN 2072 2118 LAMININ EGF-LIKE 21.
FT DOMAIN 2119 2168 LAMININ EGF-LIKE 22.
FT DOMAIN 2169 2735 DOMAIN II AND I.
FT DOMAIN 2736 2933 LAMININ G-LIKE 1.
FT DOMAIN 2934 3119 LAMININ G-LIKE 2.
FT DOMAIN 3120 3296 LAMININ G-LIKE 3.
FT DOMAIN 3297 3511 LAMININ G-LIKE 4.
FT DOMAIN 3512 3689 LAMININ G-LIKE 5.
FT DOMAIN 3690 2464 COILED COIL (POTENTIAL).
FT DOMAIN 2465 2621 COILED COIL (POTENTIAL).
FT DOMAIN 2622 2705 COILED COIL (POTENTIAL).
FT DOMAIN 2706 1725 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 1726 1841 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 305 314 BY SIMILARITY.
FT DISULFID 315 327 BY SIMILARITY.
FT DISULFID 328 338 BY SIMILARITY.
FT DISULFID 339 341 BY SIMILARITY.
FT DISULFID 342 361 BY SIMILARITY.
FT DISULFID 362 373 BY SIMILARITY.
FT DISULFID 374 398 BY SIMILARITY.
FT DISULFID 399 401 BY SIMILARITY.
FT DISULFID 402 410 BY SIMILARITY.
FT DISULFID 411 431 BY SIMILARITY.
FT DISULFID 432 445 BY SIMILARITY.
FT DISULFID 446 452 BY SIMILARITY.
FT DISULFID 453 463 BY SIMILARITY.
FT DISULFID 464 476 BY SIMILARITY.
FT DISULFID 477 500 BY SIMILARITY.
FT DISULFID 501 521 BY SIMILARITY.
FT DISULFID 522 532 BY SIMILARITY.
FT DISULFID 533 544 BY SIMILARITY.
FT DISULFID 545 559 BY SIMILARITY.
FT DISULFID 560 566 BY SIMILARITY.
FT DISULFID 567 577 BY SIMILARITY.
FT DISULFID 578 590 BY SIMILARITY.
FT DISULFID 591 605 BY SIMILARITY.
FT DISULFID 606 611 BY SIMILARITY.
FT DISULFID 612 622 BY SIMILARITY.
FT DISULFID 623 635 BY SIMILARITY.
FT DISULFID 636 650 BY SIMILARITY.
FT DISULFID 651 666 BY SIMILARITY.
FT DISULFID 667 687 BY SIMILARITY.
FT DISULFID 688 690 BY SIMILARITY.
FT DISULFID 691 680 BY SIMILARITY.
FT DISULFID 681 695 BY SIMILARITY.
FT DISULFID 696 702 BY SIMILARITY.
FT DISULFID 703 713 BY SIMILARITY.
FT DISULFID 714 726 BY SIMILARITY.
FT DISULFID 727 1443 BY SIMILARITY.
FT DISULFID 1444 1455 BY SIMILARITY.
FT DISULFID 1456 1473 BY SIMILARITY.
FT DISULFID 1474 1486 BY SIMILARITY.
FT DISULFID 1487 1533 BY SIMILARITY.
FT DISULFID 1534 1548 BY SIMILARITY.
FT DISULFID 1549 1555 BY SIMILARITY.
FT DISULFID 1556 1566 BY SIMILARITY.
FT DISULFID 1567 1579 BY SIMILARITY.
FT DISULFID 1580 1594 BY SIMILARITY.
FT DISULFID 1595 1601 BY SIMILARITY.
FT DISULFID 1602 1612 BY SIMILARITY.
FT DISULFID 1613 1630 BY SIMILARITY.
FT DISULFID 1631 1874 BY SIMILARITY.
FT DISULFID 1875 1881 BY SIMILARITY.

FT DISULFID 1884 1893 BY SIMILARITY.
FT DISULFID 1894 1912 BY SIMILARITY.
FT DISULFID 1913 1930 BY SIMILARITY.
FT DISULFID 1931 1939 BY SIMILARITY.
FT DISULFID 1940 1950 BY SIMILARITY.
FT DISULFID 1951 1968 BY SIMILARITY.
FT DISULFID 1969 1986 BY SIMILARITY.
FT DISULFID 1987 1993 BY SIMILARITY.
FT DISULFID 1994 2005 BY SIMILARITY.
FT DISULFID 2006 2022 BY SIMILARITY.
FT DISULFID 2023 2083 BY SIMILARITY.
FT DISULFID 2084 2090 BY SIMILARITY.
FT DISULFID 2091 2101 BY SIMILARITY.
FT DISULFID 2102 2116 BY SIMILARITY.
FT DISULFID 2117 2126 BY SIMILARITY.
FT DISULFID 2127 2133 BY SIMILARITY.
FT DISULFID 2134 2144 BY SIMILARITY.
FT DISULFID 2145 2166 BY SIMILARITY.
FT DISULFID 2167 2169 INTERCHAIN (PROBABLE).
FT DISULFID 2170 2172 INTERCHAIN (PROBABLE).
FT CARBOHYD 100 100 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 248 248 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 383 383 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 905 905 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 926 926 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 964 964 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1335 1335 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1534 1534 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2021 2021 N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 3.5%; Score 200; DB 1; Length 3718;
Best local Similarity 18.5%; Pred. No. 3.8e-06;
Matches 174; Conservative 82; Mismatches 279; Indels 404; Gaps 46;

QY 6 RGPVGRG-----WGRPAEAPRRGRPPWPSPAWICWALAGCOAAGDLPSS---S 54
DB 1465 RGVIGRDCSRCATGYWGFNCRP-----CDDCARLDELGTGOCICP 1506

QY 55 SRPLPP---CQEKYHFE-----YTECDSSG-SRWRAIPNSAVD---CSGLPDPVGRKE 102
DB 1507 PRTVPDCLVQCQSPGCHPLVGCECNCSGGVQELTDPTCDMDSGQCRPN-VAGR 1565

QY 103 CTFSCASGEY-----LEMKQV-----CSKCGEGTYSLS 132
DB 1566 CD-TCAPGYGYPCRCDCHEAGTMAVCDPLTGQCHCKENVQSGRCDCQVRGTFSLDA 1624

QY 133 G-----IKFDEMDLPAGFSNIATF-----MDTVVGPDSRP----- 164
DB 1625 ANPKGCTRCFCGATER--CGNSNLAHFEVDMEGWVLLSSDRQVVPHEHPRIELHLD 1682

QY 165 -----DGCNNSSWIPRGNYIESNRDCTVSLIYAVHLK-KSGYVFFYQVVDNNIFEFF 218
DB 1683 LRSVADTFSELVQAPSVLGDVRVSYGCTLHVLEHSETRQGDIFIPYESRPDV----- 1737

QY 219 IQNDQCO-----EMDTTDDKWKL-TDNGEWGSHVLMKSGTNNILYWR 260
DB 1738 LQGNQMSIAFLAYPPQGVHRGQLQVLEGNFRHLETHNPVSREELMVLVLAGLEQLIR 1797

QY 261 TTGILMGSKAVPVLVKNITIEGVAYT-----SECFCKPCGFESN 300
DB 1798 A-----LFSOTSSVSLRRVVLVASEAGRPPASNVLMCMCPANYRGDSCQECAPGYRD 1853

QY 301 KPGSF--NCQVCPRTNYSEK----GAKEICRCKDDSGSGSSECTERPPCTTKDYFIHTP 355
DB 1854 TKGLFLGRVCPQCHGSDRCLPQSGICVGCQHNTEGDQCERC--RPGFVSSDPSNPASP 1911

QY 356 CD-----EKGQTQIMYKLEPKICREDTDAIRLPPSGEKKDCPPC 396
DB 1912 CVSCPCPLAVPSNNFADGCVLRNGRTQC-----LCR-----PGVAGASCERC 1953

```
Qy 397 NPGFYNNQ-----SSSCHPC-----PPFTSDG---TKCRPC-----PAGTEPALGFE 437
Db 1954 APGFFGNPLVLGSCQPCDCSGNGDPNNIFSDCDPLTGACRGCRLRHHTGPHRCACAPGF- 2012
Qy 438 YKWN-VLPGN-MKTSFNVGNSKDCMNGHEVAGDHIQSGAGSDNDYLLNLHIFCFK 495
Db 2013 --YGNALLPGNCTRCDCSPCTETCDPSG----- 2040
Qy 496 PPTSMGTGATGSELGRITFVFETLCSADCVLFVMDINKSTNVVESWGTKEQAYTHII 555
Db 2041 -----RLCKAGVT-----GQRDCRLGFGFPCQC----- 2067
Qy 556 FKNATFTFTWAFORTNOQDNRRIINDMKIYSITATNAVDGVASSCRACALGSEQSGSS 615
Db 2068 -----GCRPCACGPAAGSE 2082
Qy 616 CVP-----CPRGHY---IEKYNCK-----ECPP----- 637
Db 2083 CHPSQGOCHCQPGTGPQCLECAPGYGLPEKGCRCRCQPRGHCDPHGTGHTCTCPPGLSGE 2142
Qy 638 --DTVLSLTHQVYKACIPCGPGSKNNO-----DHSVCYSDCFFVHEKENQILHLYDFSLS 691
Db 2143 RCDTCSQHQV-----PVPGFPGGHIHCEVCDHCV-----VLLLD--DLE 2181
Qy 692 SVGSL-----MNGPSFTSKGTGYFHFNFISLCGHEGK 723
Db 2182 RAGALLPAIREOLOGINASSAAWARLHRLNASIADLOSK 2220
RESULT 6
ZAN_MOUSE STANDARD; PRT; 5376 AA.
AC 086799; O08647;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zonadhesin precursor.
GN ZAN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=98123114; PubMed=9452463;
RA Gao Z., Garbers D.L.;
RT Species diversity in the structure of zonadhesin, a sperm-specific
RT membrane protein containing multiple cell adhesion molecule-like
RT domains."
RL J. Biol. Chem. 273:3415-3421(1998).
RN [2]
RP SEQUENCE OF 4864-5376 FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=97271566; PubMed=9126492;
RA Gao Z., Harumi T., Garbers D.L.;
RT "Chromosome localization of the mouse zonadhesin gene and the human
RT zonadhesin gene (ZAN)".
RL Genomics 41:119-122(1997).
CC -!- FUNCTION: BINDS IN A SPECIES-SPECIFIC MANNER TO THE ZONA PELLUCIDA
CC OF THE EGG. MAY BE INVOLVED IN GAMETE RECOGNITION AND/OR
CC SIGNALING.
CC -!- SUBUNIT: PROBABLY FORMS COVALENT OLIGOMERS.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN, EXCLUSIVELY ON THE
CC APICAL REGION OF THE SPERM HEAD.
CC -!- TISSUE SPECIFICITY: IN TESTIS; PRIMARILY IN HAPLOID SPERMATIDS.
CC -!- DOMAIN: THE MAM DOMAINS PROBABLY MEDIATES SPERM ADHESION TO THE
CC ZONA PELLUCIDA.
CC -!- DOMAIN: DURING SPERM MIGRATION THROUGH THE REPRODUCTIVE TRACTS,
CC THE MUCIN-LIKE DOMAIN MIGHT INHIBIT INAPPROPRIATE TRAPPING OF
CC SPERMATOZOA OR PROMOTING ADHESION TO THE OVIDUCTAL ISTHMUS.
CC -!- DOMAIN: THE VNFD DOMAIN 2 MAY MEDIATE COVALENT
CC OLIGOMERIZATION (BY SIMILARITY TO HUMAN INTESTINAL MUCIN MUC2).
```

```
CC -!- SIMILARITY: CONTAINS 3 MAM DOMAINS.
CC -!- SIMILARITY: CONTAINS 25 VNFD DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U97068; AAC26580.1; -.
DR EMBL; U83190; AAC53125.1; -.
DR MGD; MGI:106656; ZAN.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR003645; POIN.
DR InterPro; IPR000998; MAM_domain.
DR InterPro; IPR002919; TIL_Cysrich.
DR InterPro; IPR003328; TILa_Cysrich.
DR InterPro; IPR001007; VWF_C.
DR InterPro; IPR001846; VWF_D.
DR Pfam; PF000094; vwd; 4.
DR Pfam; PF00629; MAM; 3.
DR Pfam; PF01826; TIL; 25.
DR Pfam; PF02345; TILa; 25.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00274; FOLN; 11.
DR SMART; SM00137; MAM; 2.
DR SMART; SM00214; VWC; 17.
DR SMART; SM00216; VWD; 4.
DR PROSITE; PS00022; EGF_1; 1.
DR PROSITE; PS01186; EGF_2; 18.
DR PROSITE; PS00740; MAM_1; FALSE_NEG.
DR PROSITE; PS00600; MAM_2; 3.
KW Signal; Glycoprotein; Transmembrane; EGF-like domain; Cell adhesion;
KW Repeat.
FT SIGNAL 1 17
FT CHAIN 18 5376
FT ZONADHESIN.
FT EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 5311 5337
FT POTENTIAL.
FT DOMAIN 5338 5376
FT CYTOPLASMIC (POTENTIAL).
FT DOMAIN 45 210
FT MAM 1.
FT DOMAIN 215 374
FT MAM 2.
FT DOMAIN 377 542
FT MAM 3.
FT DOMAIN 547 1170
FT 80 X HEPTAPEPTIDE REPEATS (APPROXIMATE)
FT (MUCIN-LIKE DOMAIN).
FT VNFD 1 (PARTIAL).
FT VNFD 2.
FT VNFD 3.
FT VNFD 4.
FT VNFD 5 (PARTIAL).
FT VNFD 6 (PARTIAL).
FT VNFD 7 (PARTIAL).
FT VNFD 8 (PARTIAL).
FT VNFD 9 (PARTIAL).
FT VNFD 10 (PARTIAL).
FT VNFD 11 (PARTIAL).
FT VNFD 12 (PARTIAL).
FT VNFD 13 (PARTIAL).
FT VNFD 14 (PARTIAL).
FT VNFD 15 (PARTIAL).
FT VNFD 16 (PARTIAL).
FT VNFD 17 (PARTIAL).
FT VNFD 18 (PARTIAL).
FT VNFD 19 (PARTIAL).
FT VNFD 20 (PARTIAL).
FT VNFD 21 (PARTIAL).
FT VNFD 22 (PARTIAL).
FT VNFD 23 (PARTIAL).
FT VNFD 24 (PARTIAL).
FT VNFD 25.
FT EGF-LIKE.
FT DOMAIN 5259 5295
```

FT DISULFID 5263 5274 BY SIMILARITY.
 FT DISULFID 5268 5283 BY SIMILARITY.
 FT DISULFID 5285 5294 BY SIMILARITY.
 FT CARBOHYD 339 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1216 1216 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1239 1239 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1314 1314 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1814 1814 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1908 1908 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1933 1933 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2028 2028 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2111 2111 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2142 2142 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2332 2332 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2533 2533 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2575 2575 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2692 2692 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2812 2812 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3052 3052 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3065 3065 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3144 3144 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3172 3172 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3288 3288 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3292 3292 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3782 3782 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4005 4005 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4136 4136 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4243 4243 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4254 4254 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4335 4335 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4376 4376 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 4586 4586 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 5136 5136 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 5252 5252 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 5376 AA: 579908 MW: 0E44DB77DF2A2620 CRC64;

Query Match 3.4%; Score 194; DB 1; Length 5376;
 Best Local Similarity 20.5%; Pred. No. 1.7e-05;
 Matches 144; Conservative 63; Mismatches 204; Indels 290; Gaps 38;

QY 75 SSGSRWRVAIPNSAVDCSLDPVVRGKCTFSCAGSEYLEMKNOVCSKCGEGTSLGSGI 134
 DB 3256 SRGCTQSTCTCPAGATHCR-----NFKCPGSGTYCKNGDNGSSNCTEITLQCPNPS 3304

QY 135 KFDWELDPAGSFNFIATFMDTVVGPDSRPGDCGNSSWIPRGNYIESNRDCTVSLIYAV 194
 DB 3305 QFT--DCLP-----SCVPSCSNRCEVTSPSPSSCREG----- 3336

QY 195 HLKKGYPVFYQYVDNIFFEFFTNDQCQEMDTTDDKWKLTONGEWSHVMKSGT 254
 DB 3337 -LCNHGFVFE-----DKCVPRTQCGCKDARGAIIAPAG- 3368

QY 255 NILYWRRTGILMGSAKVPVLVKNITIEGVAVTSECFCKPCTFSGFNQ--VCPR 312
 DB 3369 -----KWTSGKGTQSCACV-----EGNIOQCNFQCPP 3396

QY 313 NPYSEKGAKEICRDKDSQFSGSSECTE---RPPCTTKDYFIQHTPC-----DEEG 360
 DB 3397 ETY-----CKDNSE--GSSTCTKITLQCPAHTQ-----YTSLCPSLPCLSDPEG 3439

QY 361 KTOIMYKWIPIKPI---CRE-----DLTDAI-RLPSPGSKKDCPP 395
 DB 3440 ----LCKDISPRKVPSTCKEGVCQSGYVLSNDKCVLRAECCKDAQGALIPAGKWTWTSFG 3495

QY 396 CNPG-FYNNKSSSCH--PCPPGTF-----SDGTKECR-----PCPACTEFALGFEYKWNVL 444
 DB 3496 CTQSCACMGAVOCOSSOCPCTGYCKDNEDGNSNCAKITLQCPNPS-----LFTNCL 3547

QY 445 PGNMKTSCFNVNSKCDGMNGMEVAGDIQSGAGGSDNDYLILNLHIFPKFPPTSMTGAT 504
 DB 3548 PPCL-PSCLDP-----DGL-----CKGSPKVPSTCKE 3574

QY 505 GSELGRITVEFTLCSADCVCVLFMVDINRKNSTN--VVESWGCTKEKQAYTHIFKNATFT 562
 DB 3575 G-----CICOSGYVL-----SNKCLLRNRGCKD--AHGALIPDK--- 3609

QY 563 FTWAFORTNOGDNRRFINDMVKIYSITATNAVGVASSCRACALGSESGSCVCPGPG 622
 DB 3610 -TWVSR-----GCTQSC-VCTGSGSIQCLSS--QCPG 3637

QY 623 HYI---EKETNOCKECPPTIYLSIHQVYKCAICPGCGSKNQDH-----SVCYS 670
 DB 3638 AYCKDNEDSSNCARIPPOCPANSHYT---DCFPSPSCSDPEGHCASGPRVLSTCRE 3694

QY 671 DCF-----FYHEKENQLLHYDFSNLSSVGLM-NGPSETSKG 706
 DB 3695 GCLCNPFGVLDKCVPRVECCGCKDAQGALIPSGKWTWTSFG 3735

RESULT 7
 LMB1_DROME
 ID LMB1_DROME STANDARD; PRT: 1790 AA.
 AC P11046; Q26328; Q9XZT4; Q9VLM6;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Laminin beta-1 chain precursor (Laminin B1 chain).
 GN LAMB1 OR LAMB1 OR CG7123.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 RX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Canton-S;
 RX MEDLINE=94000382; PubMed=8397815;
 RT Gow C.-H., Chang H.-Y., Lih C.-J., Chang T.-W., Hui C.-F.;
 RL "Analysis of the Drosophila gene for the laminin B1 chain.";
 RN DNA Cell Biol. 12:573-587(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=88210471; PubMed=3365769;
 RA Montell D.J., Goodman C.S.;
 RT "Drosophila substrate adhesion molecule: sequence of laminin B1 chain
 reveals domains of homology with mouse.";
 RL Cell 53:463-473(1988).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA Sutton G.G., Wortman J.K., Richards S., Ashburner M., Henderson S.N.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Zhang Q., Chen L.X.,
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
 BAILEY R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fester C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hoston D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K.A., Nusskern D.R., Pacled J.M.,
RA Palazzolo M., Pittman G.S., Pen S., Pollard J.R., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
CC -!- FUNCTION: Binding to cells via a high affinity receptor, laminin
CC is thought to mediate the attachment, migration, and organization
CC of cells into tissues during embryonic development by interacting
CC with other extracellular matrix components.
CC -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
CC different polypeptide chains (alpha, beta, gamma), which are bound
CC to each other by disulfide bonds into a cross-shaped molecule
CC comprising one long and three short arms with globules at each
CC end.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
CC COMPONENT).
CC -!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC -!- DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.
CC -!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -!- SIMILARITY: CONTAINS 12.5 LAMININ EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M95811; A019752.1; -;
CC EMBL: M19525; A028663.1; -;
CC EMBL: AE003618; AAF52563.1; -;
CC PIR: A28783; MWFFB1.
CC HSP: P02468; 1TLE.
CC FlyBase: FBgn002527; LanB1.
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR001886; LamNT.
CC InterPro: IPR002049; Laminin_EGF.
CC Pfam: PF00053; laminin_EGF; 13.
CC Pfam: PF00055; laminin_Nterm; 1.
CC PRINTS: PR00011; EGF_LAMININ.
CC ProDom: PD002082; LamNT; 1.
CC SMART: SM00180; EGF_Lam; 11.
CC SMART: SM00001; EGF_like; 1.
CC SMART: SM00136; LamNT; 1.
CC PROSITE: PS00022; EGF 1; 10.
CC PROSITE: PS01186; EGF 2; 2.
CC PROSITE: PS01248; LAMININ_TYPE_EGF; 12.
KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
FT SIGNAL 1 25
FT CHAIN 26 1790 LAMININ BETA-1 CHAIN.
FT DOMAIN 43 289 LAMININ N-TERMINAL (DOMAIN VI).
FT DOMAIN 290 356 LAMININ EGF-LIKE 1.
FT DOMAIN 357 419 LAMININ EGF-LIKE 2.
FT DOMAIN 420 479 LAMININ EGF-LIKE 3.
FT DOMAIN 480 530 LAMININ EGF-LIKE 4.
FT DOMAIN 531 561 LAMININ EGF-LIKE 5 (INCOMPLETE).
FT DOMAIN 562 789 LAMININ DOMAIN IV.
FT DOMAIN 791 838 LAMININ EGF-LIKE 6.
FT DOMAIN 839 884 LAMININ EGF-LIKE 7.

DOMAIN 885 934 LAMININ EGF-LIKE 8.
DOMAIN 935 992 LAMININ EGF-LIKE 9.
DOMAIN 993 1044 LAMININ EGF-LIKE 10.
DOMAIN 1045 1095 LAMININ EGF-LIKE 11.
DOMAIN 1096 1143 LAMININ EGF-LIKE 12.
DOMAIN 1144 1190 LAMININ EGF-LIKE 13.
DOMAIN 1191 1407 DOMAIN 11.
DOMAIN 1408 1434 DOMAIN ALPHA.
DOMAIN 1435 1790 DOMAIN I.
DOMAIN 1791 1807 COILED COIL (POTENTIAL).
DOMAIN 1808 1824 COILED COIL (POTENTIAL).
DOMAIN 1825 1841 COILED COIL (POTENTIAL).
DOMAIN 1842 1858 COILED COIL (POTENTIAL).
DOMAIN 1859 1875 BY SIMILARITY.
DOMAIN 1876 1892 BY SIMILARITY.
DOMAIN 1893 1909 BY SIMILARITY.
DOMAIN 1910 1926 BY SIMILARITY.
DOMAIN 1927 1943 BY SIMILARITY.
DOMAIN 1944 1960 BY SIMILARITY.
DOMAIN 1961 1977 BY SIMILARITY.
DOMAIN 1978 1994 BY SIMILARITY.
DOMAIN 1995 2011 BY SIMILARITY.
DOMAIN 2012 2028 BY SIMILARITY.
DOMAIN 2029 2045 BY SIMILARITY.
DOMAIN 2046 2062 BY SIMILARITY.
DOMAIN 2063 2079 BY SIMILARITY.
DOMAIN 2080 2096 BY SIMILARITY.
DOMAIN 2097 2113 BY SIMILARITY.
DOMAIN 2114 2130 BY SIMILARITY.
DOMAIN 2131 2147 BY SIMILARITY.
DOMAIN 2148 2164 BY SIMILARITY.
DOMAIN 2165 2181 BY SIMILARITY.
DOMAIN 2182 2198 BY SIMILARITY.
DOMAIN 2199 2215 BY SIMILARITY.
DOMAIN 2216 2232 BY SIMILARITY.
DOMAIN 2233 2249 BY SIMILARITY.
DOMAIN 2250 2266 BY SIMILARITY.
DOMAIN 2267 2283 BY SIMILARITY.
DOMAIN 2284 2300 BY SIMILARITY.
DOMAIN 2301 2317 BY SIMILARITY.
DOMAIN 2318 2334 BY SIMILARITY.
DOMAIN 2335 2351 BY SIMILARITY.
DOMAIN 2352 2368 BY SIMILARITY.
DOMAIN 2369 2385 BY SIMILARITY.
DOMAIN 2386 2402 BY SIMILARITY.
DOMAIN 2403 2419 BY SIMILARITY.
DOMAIN 2420 2436 BY SIMILARITY.
DOMAIN 2437 2453 BY SIMILARITY.
DOMAIN 2454 2470 BY SIMILARITY.
DOMAIN 2471 2487 BY SIMILARITY.
DOMAIN 2488 2504 BY SIMILARITY.
DOMAIN 2505 2521 BY SIMILARITY.
DOMAIN 2522 2538 BY SIMILARITY.
DOMAIN 2539 2555 BY SIMILARITY.
DOMAIN 2556 2572 BY SIMILARITY.
DOMAIN 2573 2589 BY SIMILARITY.
DOMAIN 2590 2606 BY SIMILARITY.
DOMAIN 2607 2623 BY SIMILARITY.
DOMAIN 2624 2640 BY SIMILARITY.
DOMAIN 2641 2657 BY SIMILARITY.
DOMAIN 2658 2674 BY SIMILARITY.
DOMAIN 2675 2691 BY SIMILARITY.
DOMAIN 2692 2708 BY SIMILARITY.
DOMAIN 2709 2725 BY SIMILARITY.
DOMAIN 2726 2742 BY SIMILARITY.
DOMAIN 2743 2759 BY SIMILARITY.
DOMAIN 2760 2776 BY SIMILARITY.
DOMAIN 2777 2793 BY SIMILARITY.
DOMAIN 2794 2810 BY SIMILARITY.
DOMAIN 2811 2827 BY SIMILARITY.
DOMAIN 2828 2844 BY SIMILARITY.
DOMAIN 2845 2861 BY SIMILARITY.
DOMAIN 2862 2878 BY SIMILARITY.
DOMAIN 2879 2895 BY SIMILARITY.
DOMAIN 2896 2911 BY SIMILARITY.
DOMAIN 2912 2928 BY SIMILARITY.
DOMAIN 2929 2945 BY SIMILARITY.
DOMAIN 2946 2962 BY SIMILARITY.
DOMAIN 2963 2979 BY SIMILARITY.
DOMAIN 2980 2996 BY SIMILARITY.
DOMAIN 2997 3013 BY SIMILARITY.
DOMAIN 3014 3030 BY SIMILARITY.
DOMAIN 3031 3047 BY SIMILARITY.
DOMAIN 3048 3064 BY SIMILARITY.
DOMAIN 3065 3081 BY SIMILARITY.
DOMAIN 3082 3098 BY SIMILARITY.
DOMAIN 3099 3115 BY SIMILARITY.
DOMAIN 3116 3132 BY SIMILARITY.
DOMAIN 3133 3149 BY SIMILARITY.
DOMAIN 3150 3166 BY SIMILARITY.
DOMAIN 3167 3183 BY SIMILARITY.
DOMAIN 3184 3200 BY SIMILARITY.
DOMAIN 3201 3217 BY SIMILARITY.
DOMAIN 3218 3234 BY SIMILARITY.
DOMAIN 3235 3251 BY SIMILARITY.
DOMAIN 3252 3268 BY SIMILARITY.
DOMAIN 3269 3285 BY SIMILARITY.
DOMAIN 3286 3302 BY SIMILARITY.
DOMAIN 3303 3319 BY SIMILARITY.
DOMAIN 3320 3336 BY SIMILARITY.
DOMAIN 3337 3353 BY SIMILARITY.
DOMAIN 3354 3370 BY SIMILARITY.
DOMAIN 3371 3387 BY SIMILARITY.
DOMAIN 3388 3404 BY SIMILARITY.
DOMAIN 3405 3421 BY SIMILARITY.
DOMAIN 3422 3438 BY SIMILARITY.
DOMAIN 3439 3455 BY SIMILARITY.
DOMAIN 3456 3472 BY SIMILARITY.
DOMAIN 3473 3489 BY SIMILARITY.
DOMAIN 3490 3506 BY SIMILARITY.
DOMAIN 3507 3523 BY SIMILARITY.
DOMAIN 3524 3540 BY SIMILARITY.
DOMAIN 3541 3557 BY SIMILARITY.
DOMAIN 3558 3574 BY SIMILARITY.
DOMAIN 3575 3591 BY SIMILARITY.
DOMAIN 3592 3608 BY SIMILARITY.
DOMAIN 3609 3625 BY SIMILARITY.
DOMAIN 3626 3642 BY SIMILARITY.
DOMAIN 3643 3659 BY SIMILARITY.
DOMAIN 3660 3676 BY SIMILARITY.
DOMAIN 3677 3693 BY SIMILARITY.
DOMAIN 3694 3710 BY SIMILARITY.
DOMAIN 3711 3727 BY SIMILARITY.
DOMAIN 3728 3744 BY SIMILARITY.
DOMAIN 3745 3761 BY SIMILARITY.
DOMAIN 3762 3778 BY SIMILARITY.
DOMAIN 3779 3795 BY SIMILARITY.
DOMAIN 3796 3812 BY SIMILARITY.
DOMAIN 3813 3829 BY SIMILARITY.
DOMAIN 3830 3846 BY SIMILARITY.
DOMAIN 3847 3863 BY SIMILARITY.
DOMAIN 3864 3880 BY SIMILARITY.
DOMAIN 3881 3897 BY SIMILARITY.
DOMAIN 3898 3914 BY SIMILARITY.
DOMAIN 3915 3931 BY SIMILARITY.
DOMAIN 3932 3948 BY SIMILARITY.
DOMAIN 3949 3965 BY SIMILARITY.
DOMAIN 3966 3982 BY SIMILARITY.
DOMAIN 3983 3999 BY SIMILARITY.
DOMAIN 4000 4016 BY SIMILARITY.
DOMAIN 4017 4033 BY SIMILARITY.
DOMAIN 4034 4050 BY SIMILARITY.
DOMAIN 4051 4067 BY SIMILARITY.
DOMAIN 4068 4084 BY SIMILARITY.
DOMAIN 4085 4101 BY SIMILARITY.
DOMAIN 4102 4118 BY SIMILARITY.
DOMAIN 4119 4135 BY SIMILARITY.
DOMAIN 4136 4152 BY SIMILARITY.
DOMAIN 4153 4169 BY SIMILARITY.
DOMAIN 4170 4186 BY SIMILARITY.
DOMAIN 4187 4203 BY SIMILARITY.
DOMAIN 4204 4220 BY SIMILARITY.
DOMAIN 4221 4237 BY SIMILARITY.
DOMAIN 4238 4254 BY SIMILARITY.
DOMAIN 4255 4271 BY SIMILARITY.
DOMAIN 4272 4288 BY SIMILARITY.
DOMAIN 4289 4305 BY SIMILARITY.
DOMAIN 4306 4322 BY SIMILARITY.
DOMAIN 4323 4339 BY SIMILARITY.
DOMAIN 4340 4356 BY SIMILARITY.
DOMAIN 4357 4373 BY SIMILARITY.
DOMAIN 4374 4390 BY SIMILARITY.
DOMAIN 4391 4407 BY SIMILARITY.
DOMAIN 4408 4424 BY SIMILARITY.
DOMAIN 4425 4441 BY SIMILARITY.
DOMAIN 4442 4458 BY SIMILARITY.
DOMAIN 4459 4475 BY SIMILARITY.
DOMAIN 4476 4492 BY SIMILARITY.
DOMAIN 4493 4509 BY SIMILARITY.
DOMAIN 4510 4526 BY SIMILARITY.
DOMAIN 4527 4543 BY SIMILARITY.
DOMAIN 4544 4560 BY SIMILARITY.
DOMAIN 4561 4577 BY SIMILARITY.
DOMAIN 4578 4594 BY SIMILARITY.
DOMAIN 4595 4611 BY SIMILARITY.
DOMAIN 4612 4628 BY SIMILARITY.
DOMAIN 4629 4645 BY SIMILARITY.
DOMAIN 4646 4662 BY SIMILARITY.
DOMAIN 4663 4679 BY SIMILARITY.
DOMAIN 4680 4696 BY SIMILARITY.
DOMAIN 4697 4713 BY SIMILARITY.
DOMAIN 4714 4730 BY SIMILARITY.
DOMAIN 4731 4747 BY SIMILARITY.
DOMAIN 4748 4764 BY SIMILARITY.
DOMAIN 4765 4781 BY SIMILARITY.
DOMAIN 4782 4798 BY SIMILARITY.
DOMAIN 4799 4815 BY SIMILARITY.
DOMAIN 4816 4832 BY SIMILARITY.
DOMAIN 4833 4849 BY SIMILARITY.
DOMAIN 4850 4866 BY SIMILARITY.
DOMAIN 4867 4883 BY SIMILARITY.
DOMAIN 4884 4900 BY SIMILARITY.
DOMAIN 4901 4917 BY SIMILARITY.
DOMAIN 4918 4934 BY SIMILARITY.
DOMAIN 4935 4951 BY SIMILARITY.
DOMAIN 4952 4968 BY SIMILARITY.
DOMAIN 4969 4985 BY SIMILARITY.
DOMAIN 4986 5002 BY SIMILARITY.
DOMAIN 5003 5019 BY SIMILARITY.
DOMAIN 5020 5036 BY SIMILARITY.
DOMAIN 5037 5053 BY SIMILARITY.
DOMAIN 5054 5070 BY SIMILARITY.
DOMAIN 5071 5087 BY SIMILARITY.
DOMAIN 5088 5104 BY SIMILARITY.
DOMAIN 5105 5121 BY SIMILARITY.
DOMAIN 5122 5138 BY SIMILARITY.
DOMAIN 5139 5155 BY SIMILARITY.
DOMAIN 5156 5172 BY SIMILARITY.
DOMAIN 5173 5189 BY SIMILARITY.
DOMAIN 5190 5206 BY SIMILARITY.
DOMAIN 5207 5223 BY SIMILARITY.
DOMAIN 5224 5240 BY SIMILARITY.
DOMAIN 5241 5257 BY SIMILARITY.
DOMAIN 5258 5274 BY SIMILARITY.
DOMAIN 5275 5291 BY SIMILARITY.
DOMAIN 5292 5308 BY SIMILARITY.
DOMAIN 5309 5325 BY SIMILARITY.
DOMAIN 5326 5342 BY SIMILARITY.
DOMAIN 5343 5359 BY SIMILARITY.
DOMAIN 5360 5376 BY SIMILARITY.
DOMAIN 5377 5393 BY SIMILARITY.
DOMAIN 5394 5410 BY SIMILARITY.
DOMAIN 5411 5427 BY SIMILARITY.
DOMAIN 5428 5444 BY SIMILARITY.
DOMAIN 5445 5461 BY SIMILARITY.
DOMAIN 5462 5478 BY SIMILARITY.
DOMAIN 5479 5495 BY SIMILARITY.
DOMAIN 5496 5512 BY SIMILARITY.
DOMAIN 5513 5529 BY SIMILARITY.
DOMAIN 5530 5546 BY SIMILARITY.
DOMAIN 5547 5563 BY SIMILARITY.
DOMAIN 5564 5580 BY SIMILARITY.
DOMAIN 5581 5597 BY SIMILARITY.
DOMAIN 5598 5614 BY SIMILARITY.
DOMAIN 5615 5631 BY SIMILARITY.
DOMAIN 5632 5648 BY SIMILARITY.
DOMAIN 5649 5665 BY SIMILARITY.
DOMAIN 5666 5682 BY SIMILARITY.
DOMAIN 5683 5699 BY SIMILARITY.
DOMAIN 5700 5716 BY SIMILARITY.
DOMAIN 5717 5733 BY SIMILARITY.
DOMAIN 5734 5750 BY SIMILARITY.
DOMAIN 5751 5767 BY SIMILARITY.
DOMAIN 5768 5784 BY SIMILARITY.
DOMAIN 5785 5801 BY SIMILARITY.
DOMAIN 5802 5818 BY SIMILARITY.
DOMAIN 5819 5835 BY SIMILARITY.
DOMAIN 5836 5852 BY SIMILARITY.
DOMAIN 5853 5869 BY SIMILARITY.
DOMAIN 5870 5886 BY SIMILARITY.
DOMAIN 5887 5903 BY SIMILARITY.
DOMAIN 5904 5920 BY SIMILARITY.
DOMAIN 5921 5937 BY SIMILARITY.
DOMAIN 5938 5954 BY SIMILARITY.
DOMAIN 5955 5971 BY SIMILARITY.
DOMAIN 5972 5988 BY SIMILARITY.
DOMAIN 5989 6005 BY SIMILARITY.
DOMAIN 6006 6022 BY SIMILARITY.
DOMAIN 6023 6039 BY SIMILARITY.
DOMAIN 6040 6056 BY SIMILARITY.
DOMAIN 6057 6073 BY SIMILARITY.
DOMAIN 6074 6090 BY SIMILARITY.
DOMAIN 6091 6107 BY SIMILARITY.
DOMAIN 6108 6124 BY SIMILARITY.
DOMAIN 6125 6141 BY SIMILARITY.
DOMAIN 6142 6158 BY SIMILARITY.
DOMAIN 6159 6175 BY SIMILARITY.
DOMAIN 6176 6192 BY SIMILARITY.
DOMAIN 6193 6209 BY SIMILARITY.
DOMAIN 6210 6226 BY SIMILARITY.
DOMAIN 6227 6243 BY SIMILARITY.
DOMAIN 6244 6260 BY SIMILARITY.
DOMAIN 6261 6277 BY SIMILARITY.
DOMAIN 6278 6294 BY SIMILARITY.
DOMAIN 6295 6310 BY SIMILARITY.
DOMAIN 6311 6327 BY SIMILARITY.
DOMAIN 6328 6344 BY SIMILARITY.
DOMAIN 6345 6361 BY SIMILARITY.
DOMAIN 6362 6378 BY SIMILARITY.
DOMAIN 6379 6395 BY SIMILARITY.
DOMAIN 6396 6412 BY SIMILARITY.
DOMAIN 6413 6429 BY SIMILARITY.
DOMAIN 6430 6446 BY SIMILARITY.
DOMAIN 6447 6463 BY SIMILARITY.
DOMAIN 6464 6480 BY SIMILARITY.
DOMAIN 6481 6497 BY SIMILARITY.
DOMAIN 6498 6514 BY SIMILARITY.
DOMAIN 6515 6531 BY SIMILARITY.
DOMAIN 6532 6548 BY SIMILARITY.
DOMAIN 6549 6565 BY SIMILARITY.
DOMAIN 6566 6582 BY SIMILARITY.
DOMAIN 6583 6599 BY SIMILARITY.
DOMAIN 6600 6616 BY SIMILARITY.
DOMAIN 6617 6633 BY SIMILARITY.
DOMAIN 6634 6650 BY SIMILARITY.
DOMAIN 6651 6667 BY SIMILARITY.
DOMAIN 6668 6684 BY SIMILARITY.
DOMAIN 6685 6701 BY SIMILARITY.
DOMAIN 6702 6718 BY SIMILARITY.
DOMAIN 6719 6735 BY SIMILARITY.
DOMAIN 6736 6752 BY SIMILARITY.
DOMAIN 6753 6769 BY SIMILARITY.
DOMAIN 6770 6786 BY SIMILARITY.
DOMAIN 6787 6803 BY SIMILARITY.
DOMAIN 6804 6820 BY SIMILARITY.
DOMAIN 6821 6837 BY SIMILARITY.
DOMAIN 6838 6854 BY SIMILARITY.
DOMAIN 6855 6871 BY SIMILARITY.
DOMAIN 6872 6888 BY SIMILARITY.
DOMAIN 6889 6905 BY SIMILARITY.
DOMAIN 6906 6922 BY SIMILARITY.
DOMAIN 6923 6939 BY SIMILARITY.
DOMAIN 6940 6956 BY SIMILARITY.
DOMAIN 6957 6973 BY SIMILARITY.
DOMAIN 6974 6990 BY SIMILARITY.
DOMAIN 6991 7007 BY SIMILARITY.
DOMAIN 7008 7024 BY SIMILARITY.
DOMAIN 7025 7041 BY SIMILARITY.
DOMAIN 7042 7058 BY SIMILARITY.
DOMAIN 7059 7075 BY SIMILARITY.
DOMAIN 7076 7092 BY SIMILARITY.
DOMAIN 7093 7109 BY SIMILARITY.
DOMAIN 7110 7126 BY SIMILARITY.
DOMAIN 7127 7143 BY SIMILARITY.
DOMAIN 7144 7160 BY SIMILARITY.
DOMAIN 7161 7177 BY SIMILARITY.
DOMAIN 7178 7194 BY SIMILARITY.
DOMAIN 7195 7211 BY SIMILARITY.
DOMAIN 7212 7228 BY SIMILARITY.
DOMAIN 7229 7245 BY SIMILARITY.
DOMAIN 7246 7262 BY SIMILARITY.
DOMAIN 7263 7279 BY SIMILARITY.
DOMAIN 7280 7296 BY SIMILARITY.
DOMAIN 7297 7313 BY SIMILARITY.
DOMAIN 7314 7330 BY SIMILARITY.
DOMAIN 7331 7347 BY SIMILARITY.
DOMAIN 7348 7364 BY SIMILARITY.
DOMAIN 7365 7381 BY SIMILARITY.
DOMAIN 7382 7398 BY SIMILARITY.
DOMAIN 7399 7415 BY SIMILARITY.
DOMAIN 7416 7432 BY SIMILARITY.
DOMAIN 7433 7449 BY SIMILARITY.
DOMAIN 7450 7466 BY SIMILARITY.
DOMAIN 7467 7483 BY SIMILARITY.
DOMAIN 7484 7500 BY SIMILARITY.
DOMAIN 7501 7517 BY SIMILARITY.
DOMAIN 7518 7534 BY SIMILARITY.
DOMAIN 7535 7551 BY SIMILARITY.
DOMAIN 7552 7568 BY SIMILARITY.
DOMAIN 7569 7585 BY SIMILARITY.
DOMAIN 7586 7602 BY SIMILARITY.
DOMAIN 7603 7619 BY SIMILARITY.
DOMAIN 7620 7636 BY SIMILARITY.
DOMAIN 7637 7653 BY SIMILARITY.
DOMAIN 7654 7670 BY SIMILARITY.
DOMAIN 7671 7687 BY SIMILARITY.
DOMAIN 7688 7704 BY SIMILARITY.
DOMAIN 7705 7721 BY SIMILARITY.
DOMAIN 7722 7738 BY SIMILARITY.
DOMAIN 7739 7755 BY SIMILARITY.
DOMAIN 7756 7772 BY SIMILARITY.
DOMAIN 7773 7789 BY SIMILARITY.
DOMAIN 7790 7806 BY SIMILARITY.
DOMAIN 7807 7823 BY SIMILARITY.
DOMAIN 7824 7840 BY SIMILARITY.
DOMAIN 7841 7857 BY SIMILARITY.
DOMAIN 7858 7874 BY SIMILARITY.
DOMAIN 7875 7890 BY SIMILARITY.
DOMAIN 7891 7907 BY SIMILARITY.
DOMAIN 7908 7924 BY SIMILARITY.
DOMAIN 7925 7941 BY SIMILARITY.
DOMAIN 7942 7958 BY SIMILARITY.
DOMAIN 7959 7975 BY SIMILARITY.
DOMAIN 7976 7991 BY SIMILARITY.
DOMAIN 7992 8008 BY SIMILARITY.
DOMAIN 8009 8025 BY SIMILARITY.
DOMAIN 8026 8042 BY SIMILARITY.
DOMAIN 8043 8059 BY SIMILARITY.
DOMAIN 8060 8076 BY SIMILARITY.
DOMAIN 8077 8093 BY SIMILARITY.
DOMAIN 8094 8110 BY SIMILARITY.
DOMAIN 8111 8127 BY SIMILARITY.
DOMAIN 8128 8144 BY SIMILARITY.
DOMAIN 8145 8161 BY SIMILARITY.
DOMAIN 8162 8178 BY SIMILARITY.
DOMAIN 8179 8195 BY SIMILARITY.
DOMAIN 8196 8212 BY SIMILARITY.
DOMAIN 8213 8229 BY SIMILARITY.
DOMAIN 8230 8246 BY SIMILARITY.
DOMAIN 8247 8263 BY SIMILARITY.
DOMAIN 8264 8280 BY SIMILARITY.
DOMAIN 8281 8297 BY SIMILARITY.
DOMAIN 8298 8314 BY SIMILARITY.
DOMAIN 8315 8331 BY SIMILARITY.
DOMAIN 8332 8348 BY SIMILARITY.
DOMAIN 8349 8365 BY SIMILARITY.
DOMAIN 8366 8382 BY SIMILARITY.
DOMAIN 8383 8398 BY SIMILARITY.
DOMAIN 8399 8415 BY SIMILARITY.
DOMAIN 8416 8432 BY SIMILARITY.
DOMAIN 8433 8449 BY SIMILARITY.
DOMAIN 8450 8466 BY SIMILARITY.
DOMAIN 8467 8483 BY SIMILARITY.
DOMAIN 8484 8500 BY SIMILARITY.
DOMAIN 8501 8517 BY SIMILARITY.
DOMAIN 8518 8534 BY SIMILARITY.
DOMAIN 8535 8551 BY SIMILARITY.
DOMAIN 8552 8568 BY SIMILARITY.
DOMAIN 8569 8585 BY SIMILARITY.
DOMAIN 8586 8602 BY SIMILARITY.
DOMAIN 8603 8619 BY SIMILARITY.
DOMAIN 8620 8636 BY SIMILARITY.
DOMAIN 8637 8653 BY SIMILARITY.
DOMAIN 8654 8670 BY SIMILARITY.
DOMAIN 8671 8687 BY SIMILARITY.
DOMAIN 8688 8704 BY SIMILARITY.
DOMAIN 8705 8721 BY SIMILARITY.
DOMAIN 8722 8738 BY SIMILARITY.
DOMAIN 8739 8755 BY SIMILARITY.
DOMAIN 8756 8772 BY SIMILARITY.
DOMAIN 8773 8789 BY SIMILARITY.
DOMAIN 8790 8806 BY SIMILARITY.
DOMAIN 8807 8823 BY SIMILARITY.
DOMAIN 8824 8840 BY SIMILARITY.
DOMAIN 8841 8857 BY SIMILARITY.
DOMAIN 8858 8873 BY SIMILARITY.
DOMAIN 8874 8890 BY SIMILARITY.
DOMAIN 8891 8907 BY SIMILARITY.
DOMAIN 8908 8924 BY SIMILARITY.
DOMAIN 8925 8940 BY SIMILARITY.
DOMAIN 8941 8957 BY SIMILARITY.
DOMAIN 8958 8974 BY SIMILARITY.
DOMAIN 8975 8991 BY SIMILARITY.
DOMAIN 8992 9008 BY SIMILARITY.
DOMAIN 9009 9025 BY SIMILARITY.
DOMAIN 9026 9042 BY SIMILARITY.
DOMAIN 9043 9059 BY SIMILARITY.
DOMAIN 9060 9076 BY SIMILARITY.
DOMAIN 9077 9093 BY SIMILARITY.
DOMAIN 9094 9110 BY SIMILARITY.
DOMAIN 9111 9127 BY SIMILARITY.
DOMAIN 9128 9144 BY SIMILARITY.
DOMAIN 9145 9161 BY SIMILARITY.
DOMAIN 9162 9178 BY SIMILARITY.
DOMAIN 9179 9195 BY SIMILARITY.
DOMAIN 9196 9212 BY SIMILARITY.
DOMAIN 9213 9229 BY SIMILARITY.
DOMAIN 9230 9246 BY SIMILARITY.
DOMAIN 9247 9263 BY SIMILARITY.
DOMAIN 9264 9280 BY SIMILARITY.
DOMAIN 9281 9297 BY SIMILARITY.
DOMAIN 9298 9314 BY SIMILARITY.
DOMAIN 9315 9331 BY SIMILARITY.
DOMAIN 9332 9348 BY SIMILARITY.
DOMAIN 9349 9365 BY SIMILARITY.
DOMAIN 9366 9382 BY SIMILARITY.
DOMAIN 9383 9399 BY SIMILARITY.
DOMAIN 9400 9416 BY SIMILARITY.
DOMAIN 9417 9433 BY SIMILARITY.
DOMAIN 9434 9450 BY SIMILARITY.
DOMAIN 9451 9467 BY SIMILARITY.
DOMAIN 9468 9484 BY SIMILARITY.
DOMAIN 9485 9501 BY SIMILARITY.
DOMAIN 9502 9518 BY SIMILARITY.
DOMAIN 9519 9535 BY SIMILARITY.
DOMAIN 9536 9552 BY SIMILARITY.
DOMAIN 9553 9569 BY SIMILARITY.
DOMAIN 9570 9586 BY SIMILARITY.
DOMAIN 9587 9603 BY SIMILARITY.
DOMAIN 9604 9620 BY SIMILARITY.
DOMAIN 9621 9637 BY SIMILARITY.
DOMAIN 9638 9654 BY SIMILARITY.
DOMAIN 9655 9671 BY SIMILARITY.
DOMAIN 9672 9688 BY SIMILARITY.
DOMAIN 9689 9705 BY SIMILARITY.
DOMAIN 9706 9722 BY SIMILARITY.
DOMAIN 9723 9739 BY SIMILARITY.
DOMAIN 9740 9756 BY SIMILARITY.
DOMAIN 9757 9773 BY SIMILARITY.
DOMAIN 9774 9789 BY SIMILARITY.
DOMAIN 9790 9806 BY SIMILARITY.
DOMAIN 9807 9823 BY SIMILARITY.
DOMAIN 9824 9840 BY SIMILARITY.
DOMAIN 9841 9857 BY SIMILARITY.
DOMAIN 9858 9874 BY SIMILARITY.
DOMAIN 9875 9891 BY SIMILARITY.
DOMAIN 9892 9908 BY SIMILARITY.
DOMAIN 9909 9925 BY SIMILARITY.
DOMAIN 9926 9942 BY SIMILARITY.
DOMAIN 9943 9959 BY SIMILARITY.
DOMAIN 9960 9976 BY SIMILARITY.
DOMAIN 9977 9993 BY SIMILARITY.
DOMAIN 10000 10016 BY SIMILARITY.
DOMAIN 10017 10033 BY SIMILARITY.
DOMAIN 10034 10050 BY SIMILARITY.
DOMAIN 10051 10067 BY SIMILARITY.
DOMAIN 10068 10084 BY SIMILARITY.
DOMAIN 10085 10101 BY SIMILARITY.
DOMAIN 10102 10118 BY SIMILARITY.
DOMAIN 10119 10135 BY SIMILARITY.
DOMAIN 10136 10152 BY SIMILARITY.
DOMAIN 10153 10169 BY SIMILARITY.
DOMAIN 10170 10186 BY SIMILARITY.
DOMAIN 10187 10203 BY SIMILARITY.
DOMAIN 10204 10220 BY SIMILARITY.
DOMAIN 10221 10237 BY SIMILARITY.
DOMAIN 10238 10254 BY SIMILARITY.
DOMAIN 10255 10270 BY SIMILARITY.
DOMAIN 10271 10287 BY SIMILARITY.
DOMAIN 10288 10304 BY SIMILARITY.
DOMAIN 10305 10321 BY SIMILARITY.
DOMAIN 10322 10338 BY SIMILARITY.
DOMAIN 10339 10355 BY SIMILARITY.
DOMAIN 10356 10372 BY SIMILARITY.
DOMAIN 10373 10389 BY SIMILARITY.
DOMAIN 10390 10406 BY SIMILARITY.
DOMAIN 10407 10423 BY SIMILARITY.
DOMAIN 10424 10440 BY SIMILARITY.
DOMAIN 10441 10457 BY SIMILARITY.
DOMAIN 10458 10474 BY SIMILARITY.
DOMAIN 10475 10491 BY SIMILARITY.
DOMAIN 10492 10508 BY SIMILARITY.
DOMAIN 10509 10525 BY SIMILARITY.
DOMAIN 10526 10542 BY SIMILARITY.
DOMAIN 1

Query Match 3.2%; Score 184; DB 1; Length 1790;
Best Local Similarity 18.5%; Pred. No. 2.2e-05;
Matches 259; Conservative 139; Mismatches 478; Indels 526; Gaps 69;

QY 39 LACQAAWAGDLFPSSSRP-----LPPCQ-EKDYHFEYTECDSSGRWRVAIPN- 86
DB 12 LALLSOWMD-FVDSQRPQHGRDRPKYPNNKFIKTHPCRSYCVATGNLLIGREN 70
QY 87 -SAVDCSGLPDVPR-----CKECTFCASGEYLEMKNQVCKGEGTYSLGSIKFD 137
DB 71 LTASSCGLSHPERFCLSHLQDKC-FICDTRREYKHDYKKNRIGQIIYTKPGTNIP 129
QY 138 EWDELPAFGSN-----IATFMDTVVGPDSRDPDGNSSWIPRGNYIES 181
DB 130 TWQSENGKENATIQDLAEAPHFTLITF-----TTFRPAAMYIER 172
QY 182 NRDDCIVSLIYAVHLKSGYVFEYQYVDNIFFEFFIQNDQOE-----MDTTDK 233
DB 173 SFD-----FGQTHIYR-----FAY-----DCKESFPVPTVLENITDV 207
QY 234 W-----VKLTDNGE-----WGSVMLKSGTNILYWRITGILMSKAV 271
DB 208 MCTSRYSNVEPSRNGEIVRVLPPNINVTDPYAEHVQNLKMTNLRIQMTKLKLGDNLL 267
QY 272 KPV-----VKNTIEG-----VAYTSECPKPGTFS-----NKP----- 303
DB 268 DSRLENEEKYYGISNMVVRGSCYGHASQCLPLDP-AFSQADNEDGMVHGRCCTHNT 326
QY 304 -SFNCQVC-----PRNTYSEKAGECIRCK-----DDSOFGS----- 335
DB 327 KGMNECEDEFNDLPKPAFGKTKNACKKCCNDHVSCHFEDEAVFTASGFVSGVGCND 386
QY 336 -----SECTERPPCTTKDYFO-----IHTPCD-----EAGKTOIMYKWIPI-- 373
DB 387 CLHNRGQCECMFYRDPEDQITSERVQCPCDPOGSDGDCDLSNELEEGAVAG 446
QY 374 ---CREDLTDAIRLPPSGBKCPNPGFYN---NGSSSCHPC---PPGTFSDG----- 419
DB 447 ACHKAFVTC-----RRCNQCKDGYNNLQSDNPEGCEPCTCNPLGTNNSCVMR 496
QY 420 -----TKCRPC-----PAGTE-----PALGFEYKWNVLPGNMKTSFEN 454
DB 497 TGECKCKYVTKDCNOCPETPYGLSESGSLCNCDDAGSYDNYCDVISGQCRPHM 556
QY 455 VGNKCDGMNGW-----EVAGDHIQSGAGS-----DN 482
DB 557 TGRSCSQPNFYIPLPEVHEAEVDEICISYGANGCSLVAETPDGFTGFTGRVPE 616
QY 483 DYILNL-HIPGFKP-----PTS-----MTGATGSEL 508
DB 617 SELVFTVGDIPRMPYDAVIRYQSTSRGDWENAFITLVRPDQVDPGCGCELAATSET 676
QY 509 GRITF-----VEETLCSADCVLFMDINRK-----STNVES----- 541
DB 677 -RIPFLSRLSRQVVALNEVCLAEAGKVVYFRYFERKRDHVDSPATILVDSLIPRID 735
QY 542 ----WGTX-----EQAETHIFKNATFTTFAFORTNOGQDN---RRFINDMVKIYSITA 591
DB 736 VTFPGSVLADIRKKDYEKYNCKSSLYDMYKSDPKCNLNLVSFVHDGASMCNCP 795
QY 592 TNAYDGVASSCRA-CALGSEQSGSSCVPCPG-----HYIEKET 629
DB 796 TGSLSKVCSNGGYCOCKPNVNVGROCDQAPGTGYGPEGCKACDNCSTGSKDKYCDLIT 855
QY 630 NOCKECPPTILSIHGVYKEACIPCGSKKNODHSVCYS-----DCRF 674
DB 856 GOC-QCVNPT-----YGRE-CNQCQPYWNPPECRVCQCNGHAATCDPIQGTICD-- 904
QY 675 YHEKENQILHYDFSNLSSVGLMWG-----PSFTSK-----GTKYF 710
DB 905 -----QDSTTGSCDCLDGYGNPLFGSEIGRCRCPETVASGLAHAGDCSLD 954

QY 711 HFFNISLC-----GHEGKKMALCTNNITDFTVKEIVAGSDDYTNLGVAFVCOSTIIPSES- 765
DB 955 TRNNMLCHQEGYSGSRCEICADNF-----FGPNDNGGTCKSCGCSNNVDLYDTG 1005
QY 766 -----KGFRAALSSOSIILADTFIGVTIVETTLK-----NINIKEDMFPVP 805
DB 1006 NCDROTGACLKCLYQYTGDCHELCCKGFFGDALQOQCCECEDFLGTNNNTIAHCDRTGQ 1065
QY 806 TSQIPDV-----HFFYKSTATTSCINGRSTAVKMRNPTKSGAGVISVPSKCP 854
DB 1066 CPCLPNVQGVRCDOCAENHMKIASGEGCESCNDPIGALHEQCQ---SYTG-----OCQCK 1118
QY 855 AG---TCDCCTTYELWESAEACPLCTEHDHFIEIGACKKGFQETL-----YVWNEPK 903
DB 1119 PGFGGRACQOQAHYWGPNNEKQPCCEQOFGAADFQCDRETGNCVCHGIGYKCN-- 1176
QY 904 WCIGKISLPEKKIATC-ETVDFW-LKVGAGVGAFATVLL-----VALTCYFWKK 950
DB 1177 -CARGVIGQFPHCPCGECFNNWDLILSALEDDATTILRAKEIKQVGATGAYTSESEL 1235
QY 951 NQKLEY-----KYSKLVTMTNSKECELPAADESCAIMEGDEEVEEYVYVSNKQS 997
DB 1236 DKKLQHIRNLQNTSVSLVDIEKLDYETQSLRDLQOASHG-RLSETEQNLDIY--NSLS 1292
QY 998 LLG-KLKSLATREKEDHFESVQ 1018
DB 1293 LSGVELES-----QNHSLRVQ 1309
RESULT 8
TS11_GIALA STANDARD; PRT; 667 AA.
AC Q03185;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Major surface trophozoite antigen 11 precursor.
OS TS11.
GN Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardinales; Giardia.
OX NCBI_TaxId=5741;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Isolate AD-1;
RX MEDLINE=93241215; PubMed=8479449;
RA By P.L., Khanna K., Manning P.A., Mayrhofer G.;
RT "A gene encoding a 69-kilodalton major surface protein of Giardia intestinalis trophozoites."
RL Mol. Biochem. Parasitol. 58:247-258(1993).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ANCHORED TO THE PLASMA MEMBRANE.
CC -!- DOMAIN: CONTAINS 27 REPEATS OF A CXXC MOTIF.
CC -!- SIMILARITY: BELONGS TO THE GIARDIA VARIANT SURFACE PROTEIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL: M95814; AAA02687.1;
DR PIR: A48579; A48579.
DR HSP: P02468; IKLO.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR005127; Giardia_vsp.
DR Pfam: PF03302; VSP; 1.
DR SMART: SM00181; EGF; 3.
DR SMART: SM00261; FU; 5.
DR Antigen; Repeat; Transmembrane; Signal.
KW SIGNAL 1 17
FT

```
FT CHAIN 18 667 MAJOR SURFACE TROPHOZOITE ANTIGEN 11.
FT DOMAIN 18 633 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 634 662 POTENTIAL.
FT DOMAIN 663 667 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 591 591 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 630 630 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 667 AA: 68475 MW: 1DD95727032328D CRC64:

Query Match 3.2%; Score 182; DB 1: Length 667;
Best Local Similarity 19.9%; Pred. No. 8.6e-06;
Matches 146; Conservative 58; Mismatches 236; Indels 292; Gaps 40:

Qy 103 CTFSCASGEYLEMKNQVCKGEG-TYISLGSGIKFDEWDELPAFESNIATFMDTVVGPSD 161
Db 141 CSAPKAPGEDNTPKAATCTKCAAGFLHTPSGLSSCE-ETCPPEGY-----FGHTATAESK 194
Qy 162 SRPDG-NNSSWTPRGNYTESNRDCTVSLIYAVHLKSGYVFEYQYVDNIFFEFTQ 220
Db 195 KTKCKSCTGGSSEAPNVKGI---GDCLKCM-----YNEASGNLT----- 229
Qy 221 NDOCOEMDTTDDKWKLTONGEWSHVMKSGTNILYWRITG-----ILMGSKA 270
Db 230 --TCEKSAQKPSLDKTCNDG-----TGQCAFCSGSGGCGDCSGFILDGQNC 279
Qy 271 VKPVLVKNITIEGVAYTSECFPCPKGTESNKPQSFNCQVCPNRTYSEKAKE-CIRCKDD 329
Db 280 VK-----SDC-----KTENCKACT---NPKAANEVCTECIST 308
Qy 330 SFGSGSECTERPPCTTKDYFIHTPCDEBQKTOIMYKWIPIEKICREDLTDAILRPPSGE 389
Db 309 HILTPTSQCVQY--COALGNYYAGTNADNK-----KACKE--CTVANCKTCTND 352
Qy 390 KKDPCPCNPGFVNGSS-----SCHPCPPGTFSDGTEKCRPCPA-----GTEPAL 434
Db 353 QGQCQTCNDGFKYNGDACPCHESCKTCSAGTASD-----CTECPGCKALKYNGDGTGKTC 408
Qy 435 G-----FEYKWNVLPGN-----MKTSCFNVG 456
Db 409 GEGCTTGQSGGACKTCGLTIDGASYCSECDTONEYPONGICTSTTARTVATCKNS--NVA 466
Qy 457 NSKCDG-----MNG--NEVAGDHIQSGAGSDNDYLLNLHLIPGFKPPTSMGTATGSE 507
Db 467 NGTCSCTNGFLRMNGGCVETTKFPGKSVCEGANADADCKAPVPQYK-----VE 516
Qy 508 LGRITVEFTLCSADCVLFMVDINKSTNVVESWGTEKQKQAYTHIIFKNATFTFTWAF 567
Db 517 AGKL-----VMCSKGC-----DTCSDATTC-----TKCGDGYTKI----- 546
Qy 568 QRTNQGDNRFRINDVMKYISITATNAVGVASSCRACALGSEQSGSSCVPCPPGHY--- 624
Db 547 -----EN-----SQTCTKC-----DSSCETCT---GAATTCKVCATGYKYKA 580
Qy 625 IEKETNOCKECPDPTVLSITHOVYGEACIPCGPGSKNNODHSVCYSDCFYHEKENQILH 684
Db 581 LGEST--CTSCENDS-----NGVIGVKGCLNCAPPS----- 609
Qy 685 YDFSNLSSVGLSNGPSPFTSKGTFKFFHFNISLCGHEGKKMALCTNNITDFTVKEIVAGS 744
Db 610 -----SSTGSVL--CYLMKGEN-----TGGSVNKSGLSLTCAGISVAVVV-- 649
Qy 745 DDVTNLVGAFCV 756
Db 650 --VAGLVG-FLC 658
```

RESULT 9

```
LM2_CABEL STANDARD: PRT: 3672 AA.
ID LM2_CABEL
AC Q21313;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Laminin-like protein K08C7.3 precursor.
```

```
GN K08C7.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCRL_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Berks M.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -!- SIMILARITY: CONTAINS 21.5 LAMININ EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.
CC -!- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z70286; CAA94293.1; .
CC HSSP; P02468; 1KLO.
CC WormPep; K08C7.3; CE06136.
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR001886; LamNT.
CC InterPro; IPR000034; Laminin_B.
CC InterPro; IPR002049; Laminin_EGF.
CC InterPro; IPR001791; Laminin_G.
CC Pfam; PF00052; Laminin_B; 1.
CC Pfam; PF00053; Laminin_EGF; 21.
CC Pfam; PF00054; Laminin_G; 5.
CC Pfam; PF00055; Laminin_Nterm; 1.
CC PRINTS; PR00011; EGF_LAMININ.
CC ProDom; PD002082; LamNT; 1.
CC ProDom; PD003031; Laminin_B; 1.
CC SMART; SM00180; EGF_Lam; 21.
CC SMART; SM00281; LamB; 1.
CC SMART; SM00282; LamG; 5.
CC SMART; SM00136; LamNT; 1.
CC PROSITE; PS00022; EGF_1; 19.
CC PROSITE; PS01186; EGF_2; 4.
CC PROSITE; PS01248; LAMININ_TYPE_EGF; 21.
CC PROSITE; PS50025; LAM_G_DOMAIN; 5.
KW Hypothetical protein; Laminin EGF-like domain; Signal; Repeat.
FT SIGNAL 1 27
FT CHAIN 28 3672 LAMININ-LIKE PROTEIN K08C7.3.
FT DOMAIN 28 297 LAMININ N-TERMINAL (DOMAIN VI).
FT DOMAIN 298 356 LAMININ EGF-LIKE 1.
FT DOMAIN 357 426 LAMININ EGF-LIKE 2.
FT DOMAIN 427 471 LAMININ EGF-LIKE 3.
FT DOMAIN 472 518 LAMININ EGF-LIKE 4.
FT DOMAIN 519 563 LAMININ EGF-LIKE 5.
FT DOMAIN 564 609 LAMININ EGF-LIKE 6.
FT DOMAIN 610 655 LAMININ EGF-LIKE 7.
FT DOMAIN 656 700 LAMININ EGF-LIKE 8.
FT DOMAIN 701 755 LAMININ EGF-LIKE 9.
FT DOMAIN 756 808 LAMININ EGF-LIKE 10.
FT DOMAIN 809 839 LAMININ EGF-LIKE 11 (INCOMPLETE).
FT DOMAIN 1415 1460 LAMININ EGF-LIKE 12.
FT DOMAIN 1461 1505 LAMININ EGF-LIKE 13.
FT DOMAIN 1506 1553 LAMININ EGF-LIKE 14.
FT DOMAIN 1554 1604 LAMININ EGF-LIKE 15.
FT DOMAIN 1605 1614 LAMININ EGF-LIKE 16 (N-TERMINAL).
FT DOMAIN 1615 1796 LAMININ DOMAIN IV.
FT DOMAIN 1797 1829 LAMININ EGF-LIKE 17.
FT DOMAIN 1830 1879 LAMININ EGF-LIKE 18.
FT DOMAIN 1880 1936 LAMININ EGF-LIKE 19.
FT DOMAIN 1937 1989 LAMININ EGF-LIKE 20.
FT DOMAIN 1990 2036 LAMININ EGF-LIKE 21.
FT DOMAIN 2037 2083 LAMININ EGF-LIKE 21.
```

FT	DISULFID	1973	1987	BY SIMILARITY.
FT	DISULFID	1990	2000	BY SIMILARITY.
FT	DISULFID	1992	2007	BY SIMILARITY.
FT	DISULFID	2009	2018	BY SIMILARITY.
FT	DISULFID	2021	2034	BY SIMILARITY.
FT	DISULFID	2037	2048	BY SIMILARITY.
FT	DISULFID	2039	2055	BY SIMILARITY.
FT	DISULFID	2057	2066	BY SIMILARITY.
FT	DISULFID	2069	2081	BY SIMILARITY.
FT	DISULFID	2084	2096	BY SIMILARITY.
FT	DISULFID	2086	2103	BY SIMILARITY.
FT	DISULFID	2105	2114	BY SIMILARITY.
FT	DISULFID	2117	2129	BY SIMILARITY.
FT	CARBOHYD	121	121	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	140	140	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	249	249	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	351	351	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	477	477	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	511	511	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	530	530	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	634	634	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	761	761	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1014	1014	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1341	1341	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1705	1705	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1756	1756	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1868	1868	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1944	1944	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1986	1986	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2002	2002	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2159	2159	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2207	2207	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2231	2231	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2235	2235	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2401	2401	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2421	2421	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2487	2487	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	2821	2821	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3087	3087	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3242	3242	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	3541	3541	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	SEQUENCE	3672	AA; 404223 MW; 28E262DB5FF14BFA	CRC64;
Query Match 3.1%; Score 178.5; DB 1; Length 3672;				
Best Local Similarity 22.6%; Pred. No. 0.00015;				
Matches 116; Conservative 54; Mismatches 203; Indels 141; Gaps				
QY	279	ITIGVAYTSECPCPKP--GTFENKPGSF--NQVCPRNITYSEKGAKECIRCKDSDQSG	334	
Db	1413	LSCDCAQGSSESFOCEYGGQCKPKGVIGRRERCAPGYN---FPECIKC-----QCNA	1465	
QY	335	SSECTERP-PCCTTKDYFIHTPCDESGKQTQIMYKWTEPKI-----CREDLTDAIRL---	384	
Db	1466	QQQCDERTGGCFPPHVEGOT-CDRCVSNAGFY---DPLIGCQKCGCHPQSGEGNVLVCD	1521	
QY	385	PPSGE-----KKDCPPCNPGFYNNNGSSSCHPCP---PGT---FSDGTK-----	421	
Db	1522	PESGQCLRESMGRQCDRLAGFY-GPPHYGCSNCRAGTTEELICDATNAQCKCKENV	1579	
QY	422	---BCRPCPAGTEPALGFYKWNVLPGNMKTSFCFNVGNSKCDGMNGWEVAGDHITQSGAG	478	
Db	1580	YGGRCRCAKAGT-----FDLSAENPL-GCVNCFPCFVTDSCRSSMYPVTIMSVDMSSFLT	1633	
QY	479	GSDN-----DYLIILNLHHPG-----FXPPSTMTGATGSELGRITFFVFETL-----	519	
Db	1634	TDDNGMVDNKDDTVITVSETPSNSVYFNVPTEKKDYTTSYGLKTLFKLSTVPRGGRKSM	1693	
QY	520	SADCVLFEMVDINRKNSTNV-VESWGG---TKEQAYTHIIKFNATFTTTFWAFORTNOQOD	575	
Db	1694	NAD-----ADVRLTGANMTIEWASEQTPNEEQPT-VKCLVLPENFLTABGKTVTREETREE	1746	
QY	576	NRRFINDMVKI-----YSTATNAVDGVY---ASSCRACALGSEQSGS	614	

Db 1747 LMKVLHSLQNTLKASYFDHPKSTLYEFLGSEBPNVSDSVIKASSVQCCPAPYTG 1806
 Qy 615 SCVCPGPHYIEKETNQCCEPDYSLTHQVYKRE---ACIPC---GPGSKNODHSVC 668
 Db 1807 SCLCASYG-----HRVQSGFLCACYPCCEGNHSGATCDPTGTGIC 1846
 Qy 669 YSDCFYHEKNEQILHYDFSNISSVGLMNGPSF 702
 Db 1847 -TDC-----EHNTNGDCEPCNGHYGNATNGSPY 1875

RESULT 10
 LMGL_MOUSE
 ID LMGL_MOUSE STANDARD: PRT: 1607 AA.
 AC P02468;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Laminin gamma-1 chain precursor (Laminin B2 chain).
 GN LAMC1 OR LAMB-1 OR LAMB-2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88059118; PubMed=3680290;
 RA Sasaki M., Yamada Y.;
 RT "The laminin B2 chain has a multidomain structure homologous to the
 B1 chain.";
 RL J. Biol. Chem. 262:17111-17117(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89000737; PubMed=3167041;
 RA Durkin M.E., Bartos B.B., Liu S.-H., Phillips S.L., Chung A.F.;
 RT "Primary structure of the mouse laminin B2 chain and comparison with
 laminin B1.";
 RL Biochemistry 27:5198-5204(1988).
 RN [3]
 RP SEQUENCE OF 1-239 FROM N.A.
 RX MEDLINE=88228071; PubMed=2836421;
 RA Ogawa K., Burelo P.D., Sasaki M., Yamada Y.;
 RT "The laminin B2 chain promoter contains unique repeat sequences and
 is active in transient transfection.";
 RL J. Biol. Chem. 263:8384-8389(1988).
 RN [4]
 RP SEQUENCE OF 1391-1607 FROM N.A.
 RX MEDLINE=85051302; PubMed=6209134;
 RA Barlow D.P., Green N.M., Kurkinen M., Hogan B.L.M.;
 RT "Sequencing of laminin B chain cDNAs reveals C-terminal regions of
 coiled-coil alpha-helix.";
 RL EMBO J. 3:2355-2362(1984).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 771-932.
 RX MEDLINE=96196434; PubMed=8648630;
 RA Steinfeld J., Mayer U., Timpi R., Huber R.;
 RT "Crystal structure of three consecutive laminin-type epidermal growth
 factor-like (LE) modules of laminin gamma1 chain harboring the
 nidogen binding site.";
 RL J. Mol. Biol. 257:644-657(1996).
 RN [6]
 RP STRUCTURE BY NMR OF 824-881.
 RX MEDLINE=96196435; PubMed=8648631;
 RA Baumgartner R., Csisz M., Mayer U., Poeschl E., Huber R.,
 RA Timpi R., Holak T.A.;
 RT "Structure of the nidogen binding LE module of the laminin gamma1
 chain in solution.";
 RL J. Mol. Biol. 257:658-668(1996).
 CC -1- FUNCTION: Binding to cells via a high affinity receptor, laminin
 is thought to mediate the attachment, migration, and organization
 of cells into tissues during embryonic development by interacting
 with other extracellular matrix components.
 CC -1- SUBUNIT: Laminin is a complex glycoprotein, consisting of three

CC different polypeptide chains (alpha, beta, gamma), which are bound
 CC to each other by disulfide bonds into a cross-shaped molecule
 CC comprising one long and three short arms with globules at each
 CC end.
 CC THE GAMMA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ).
 CC LAMININ-2 (MEROSIN), LAMININ-3 (S-LAMININ), LAMININ-4 (S-MEROSIN),
 CC LAMININ-6 (K-LAMININ) AND LAMININ-7 (KS-LAMININ).
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
 CC COMPONENT).
 CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
 CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
 CC -1- DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.
 CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
 CC -1- SIMILARITY: CONTAINS 11 LAMININ EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X05211; CAA28838.1; -;
 CC EMBL; J03484; AAA39405.1; -;
 CC EMBL; J02930; AAA39408.1; -;
 CC EMBL; J03749; AAA39409.1; -;
 CC PIR; A28469; MMSB2.
 CC PDB; 1KLO; 20-AUG-97.
 CC PDB; 1TLE; 12-FEB-97.
 CC MGD; MGI:99914; Lamc1.
 CC InterPro; IPR004089; Chmtaxis_transd.
 CC InterPro; IPR000561; EGF-like.
 CC InterPro; IPR001866; LamNT.
 CC InterPro; IPR000034; Laminin_B.
 CC InterPro; IPR002049; Laminin_EGF.
 CC Pfam; PF00052; Laminin_B; 1.
 CC Pfam; PF00053; Laminin_EGF; 10.
 CC Pfam; PF00055; Laminin_Nterm; 1.
 CC PRINTS; PR00011; EGF_LAMININ.
 CC PRODOM; PD002082; LamNT; 1.
 CC PRODOM; PD003031; Laminin_B; 1.
 CC SMART; SM00180; EGF_Lam; 9.
 CC SMART; SM00001; EGF_like; 1.
 CC SMART; SM00281; LamB; 1.
 CC SMART; SM00136; LamNT; 1.
 CC PROSITE; PS00022; EGF_1; 8.
 CC PROSITE; PS01186; EGF_2; 2.
 CC PROSITE; PS01248; LAMININ_TYPE_EGF; 10.
 CC KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
 KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal; 3D-structure.
 CC
 FT SIGNAL 1 33
 FT CHAIN 34 1607 LAMININ GAMMA-1 CHAIN
 FT DOMAIN 34 283 LAMININ N-TERMINAL (DOMAIN VI).
 FT DOMAIN 284 339 LAMININ EGF-LIKE 1.
 FT DOMAIN 340 395 LAMININ EGF-LIKE 2.
 FT DOMAIN 396 442 LAMININ EGF-LIKE 3.
 FT DOMAIN 443 492 LAMININ EGF-LIKE 4.
 FT DOMAIN 493 502 LAMININ EGF-LIKE 5 (N-TERMINAL).
 FT DOMAIN 503 687 LAMININ DOMAIN IV.
 FT DOMAIN 688 721 LAMININ EGF-LIKE 5 (C-TERMINAL).
 FT DOMAIN 722 770 LAMININ EGF-LIKE 6.
 FT DOMAIN 771 825 LAMININ EGF-LIKE 7.
 FT DOMAIN 826 881 LAMININ EGF-LIKE 8 (NIDOGEN-BINDING).
 FT DOMAIN 882 932 LAMININ EGF-LIKE 9.
 FT DOMAIN 933 980 LAMININ EGF-LIKE 10.
 FT DOMAIN 981 1038 LAMININ EGF-LIKE 11.
 FT DOMAIN 1039 1097 DOMAIN II AND I.
 FT DOMAIN 1098 1594 COILED COIL (POTENTIAL).
 FT DISULFID 340 349 BY SIMILARITY.
 FT DISULFID 342 365 BY SIMILARITY.

FBN2_MOUSE
ID FBN2_MOUSE STANDARD; PRT; 2907 AA.
AC Q61555; Q63957;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Fibrillin 2 precursor.
DN FBN2 OR FBN-2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=95263670; PubMed=7744963;
RA Zhang H., Hu W., Ramirez F.;
RT "Developmental expression of fibrillin genes suggests heterogeneity
of extracellular microfibrils";
RL J Cell Biol. 129:1165-1176(1995).
[2]
RN SEQUENCE OF 210-317 FROM N.A.
RP MEDLINE=94140368; PubMed=8307578;
RA Li X., Pereira L., Zhang H., Sanguineti C., Ramirez F., Bonadio J.,
RT Francke U.;
RT "Fibrillin genes map to regions of conserved mouse/human synteny on
mouse chromosomes 2 and 18";
RL Genomics 18:667-672(1993).
CC -!- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS
THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE
LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.
CC -!- SIMILARITY: CONTAINS 47 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; L39790; AAA74908.1; .
DR EMBL; S69359; AAC60685.1; .
DR HSSP; P35555; IEMN.
DR MGD; MGI:95490; Fbn2.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF-II.
DR InterPro; IPR002212; Fibril-assoc.
DR Pfam; PF00008; EGF_46.
DR Pfam; PF00683; TB; 9.
DR PRINTS; PR00010; EGFBLOOD.
DR SMART; SM00179; EGF_CA; 43.
DR SMART; SM00001; EGF_like; 3.
DR PROSITE; PS00010; ASX_HYDROXYL; 43.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 36.
DR PROSITE; PS01187; EGF_CA; 43.
KW Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;
Repeat; Signal; Multigene family.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 2907 FIBRILLIN 2.
FT DOMAIN 111 142 EGF-LIKE 1.
FT DOMAIN 145 176 EGF-LIKE 2.
FT DOMAIN 175 208 EGF-LIKE 3.
FT DOMAIN 216 248 EGF-LIKE 4.
FT DOMAIN 276 317 EGF-LIKE 5, CALCIUM-BINDING.
FT DOMAIN 318 359 EGF-LIKE 6, CALCIUM-BINDING.
FT REPEAT 360 426 TGFBP 1.
FT DOMAIN 427 527 EGF-LIKE 7.
FT DOMAIN 528 567 EGF-LIKE 8, CALCIUM-BINDING.
FT DOMAIN 568 609 EGF-LIKE 9, CALCIUM-BINDING.
FT DOMAIN 610 650

FT	DISULFID	677	690	BY SIMILARITY.
FT	DISULFID	765	777	BY SIMILARITY.
FT	DISULFID	772	786	BY SIMILARITY.
FT	DISULFID	788	801	BY SIMILARITY.
FT	DISULFID	807	819	BY SIMILARITY.
FT	DISULFID	814	828	BY SIMILARITY.
FT	DISULFID	830	843	BY SIMILARITY.
FT	DISULFID	849	859	BY SIMILARITY.
FT	DISULFID	854	868	BY SIMILARITY.
FT	DISULFID	870	883	BY SIMILARITY.
FT	DISULFID	952	964	BY SIMILARITY.
FT	DISULFID	959	973	BY SIMILARITY.
FT	DISULFID	975	988	BY SIMILARITY.
FT	DISULFID	1070	1082	BY SIMILARITY.
FT	DISULFID	1077	1091	BY SIMILARITY.
FT	DISULFID	1093	1106	BY SIMILARITY.
FT	DISULFID	1112	1124	BY SIMILARITY.
FT	DISULFID	1119	1133	BY SIMILARITY.
FT	DISULFID	1135	1149	BY SIMILARITY.
FT	DISULFID	1155	1167	BY SIMILARITY.
FT	DISULFID	1162	1176	BY SIMILARITY.
FT	DISULFID	1178	1191	BY SIMILARITY.
FT	DISULFID	1197	1209	BY SIMILARITY.
FT	DISULFID	1204	1218	BY SIMILARITY.
FT	DISULFID	1220	1233	BY SIMILARITY.
FT	DISULFID	1239	1250	BY SIMILARITY.
FT	DISULFID	1246	1259	BY SIMILARITY.
FT	DISULFID	1261	1274	BY SIMILARITY.
FT	DISULFID	1280	1292	BY SIMILARITY.
FT	DISULFID	1287	1301	BY SIMILARITY.
FT	DISULFID	1303	1316	BY SIMILARITY.
FT	DISULFID	1322	1334	BY SIMILARITY.
FT	DISULFID	1329	1343	BY SIMILARITY.
FT	DISULFID	1345	1358	BY SIMILARITY.
FT	DISULFID	1364	1377	BY SIMILARITY.
FT	DISULFID	1371	1386	BY SIMILARITY.
FT	DISULFID	1388	1399	BY SIMILARITY.
FT	DISULFID	1405	1418	BY SIMILARITY.
FT	DISULFID	1412	1427	BY SIMILARITY.
FT	DISULFID	1429	1440	BY SIMILARITY.
FT	DISULFID	1446	1458	BY SIMILARITY.
FT	DISULFID	1453	1467	BY SIMILARITY.
FT	DISULFID	1469	1482	BY SIMILARITY.
FT	DISULFID	1488	1499	BY SIMILARITY.
FT	DISULFID	1494	1508	BY SIMILARITY.
FT	DISULFID	1510	1523	BY SIMILARITY.
FT	DISULFID	1529	1540	BY SIMILARITY.
FT	DISULFID	1535	1549	BY SIMILARITY.
FT	DISULFID	1551	1564	BY SIMILARITY.
FT	DISULFID	1647	1659	BY SIMILARITY.
FT	DISULFID	1654	1668	BY SIMILARITY.
FT	DISULFID	1670	1683	BY SIMILARITY.
FT	DISULFID	1689	1701	BY SIMILARITY.
FT	DISULFID	1696	1710	BY SIMILARITY.
FT	DISULFID	1712	1725	BY SIMILARITY.

Query Match

Best Local Similarity

Matches 224; Conservative 101; Mismatches 375; Indels 417; Gaps

3.1%; Score 177.5; DB 1; Length 2907;

20.1%; Pred. No. 0.00013;

QY	41	GCQAAWAGDL-----PSSSRPLPPCQEKDYHFFETCDSSGSRWR--VAIPNSAVDC	91
Db	17	GCVALMAQGTGDPQPPPKTLWPQPPOVRPAVAGSEGGFMGPYRDGAVRASRRR	76
QY	92	SGLPDPVRGKE-C-----TFSCASGEYLEMKNO---VC-SKCGEGTYS-----	129
Db	77	RGOQEILRGPVNCGRSFHSYCCPGWTLPGNGNCIYPICRNSCGDGFCSRPNMCTCSSQ	136
QY	130	-----LGSGIRKFEWDELPAQFSNIATFMDTVVGPSDSRP---DGCNN	169
Db	137	ISPTCGRKSTQOCVSRVCMNGTCAHDHCOCQGY-----ICGYCCQPVCMNGCON	186
QY	170	SSWTPRGNYESNRDDCTSLIYAVHLKKSQYVF-----FEYQVDVNNIEFFFIQNDCCQ	225

OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Albus A.M., Burgess B., Champlaud M.-F., Koch M., Olson P.;
 RT "Mouse laminin 12 gamma 3 chain";
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 1-1526 FROM N.A.
 RA MEDLINE=99253969; PubMed=10318827;
 RA Iivanainen A., Morita T., Tryggvason K.;
 RT "Molecular cloning and tissue-specific expression of a novel murine
 laminin gamma3 chain";
 RL J. Biol. Chem. 274:14107-14111(1999).
 CC -!- FUNCTION: Binding to cells via a high affinity receptor, laminin
 is thought to mediate the attachment, migration, and organization
 of cells into tissues during embryonic development by interacting
 with other extracellular matrix components.
 CC -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
 different polypeptide chains (alpha, beta, gamma), which are bound
 to each other by disulfide bonds into a cross-shaped molecule
 comprising one long and three short arms with globules at each
 end.
 CC THE GAMMA-3 CHAIN IS A SUBUNIT OF LAMININ-12.
 CC -!- SUBCELLULAR LOCATION: Extracellular.
 CC -!- TISSUE SPECIFICITY: Strongly expressed in capillaries and
 arterioles of kidney as well as in interstitial Leydig cells of
 testis.
 CC -!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
 WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
 CC -!- DOMAIN: DOMAIN IV IS GLOBULAR.
 CC -!- SIMILARITY: CONTAINS 11 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
 CC -!- SIMILARITY: CONTAINS 11 LAMININ EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation.
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF083372; AAF08983.1; -;
 DR EMBL; AF079520; AAD29851.1; -;
 DR HSSP; P02468; 1KLO.
 DR MGD; MGI:1344394; Lamc3.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001886; Lamnt.
 DR InterPro; IPR000034; Laminin_B.
 DR InterPro; IPR002049; Laminin_EGF.
 DR Pfam; PF00052; laminin_B; 1.
 DR Pfam; PF00053; laminin_EGF; 10.
 DR Pfam; PF00055; laminin_Nterm; 1.
 DR PRINTS; PR00011; EGF_LAMININ.
 DR ProDom; PD002082; Lamnt; 1.
 DR SMART; SM00180; EGF_Lam; 10.
 DR SMART; SM00281; LamB; 1.
 DR SMART; SM00136; Lamnt; 1.
 DR PROSITE; PS00022; EGF_1; UNKNOWN_8.
 DR PROSITE; PS01186; EGF_2; 2.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
 DR Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
 KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
 FT SIGNAL 1 28
 FT CHAIN 29 1581
 FT DOMAIN 29 279 LAMININ GAMMA-3 CHAIN.
 FT DOMAIN 29 279 LAMININ N-TERMINAL (DOMAIN VI).
 FT DOMAIN 280 335 LAMININ EGF-LIKE 1.
 FT DOMAIN 336 391 LAMININ EGF-LIKE 2.
 FT DOMAIN 392 438 LAMININ EGF-LIKE 3.
 FT DOMAIN 439 488 LAMININ EGF-LIKE 4.

FT	DOMAIN	489	498	507	516	525	534	543	552	561	570	579	588	597	606	615	624	633	642	651	660	669	678	687	696	705	714	723	732	741	750	759	768	777	786	795	804	813	822	831	840	849	858	867	876	885	894	903	912	921	930	939	948	957	966	975	984	993	1002	1011	1020	1029	1038	1047	1056	1065	1074	1083	1092	1101	1110	1119	1128	1137	1146	1155	1164	1173	1182	1191	1200	1209	1218	1227	1236	1245	1254	1263	1272	1281	1290	1299	1308	1317	1326	1335	1344	1353	1362	1371	1380	1389	1398	1407	1416	1425	1434	1443	1452	1461	1470	1479	1488	1497	1506	1515	1524	1533	1542	1551	1560	1569	1578	1587	1596	1605	1614	1623	1632	1641	1650	1659	1668	1677	1686	1695	1704	1713	1722	1731	1740	1749	1758	1767	1776	1785	1794	1803	1812	1821	1830	1839	1848	1857	1866	1875	1884	1893	1902	1911	1920	1929	1938	1947	1956	1965	1974	1983	1992	2001	2010	2019	2028	2037	2046	2055	2064	2073	2082	2091	2100	2109	2118	2127	2136	2145	2154	2163	2172	2181	2190	2199	2208	2217	2226	2235	2244	2253	2262	2271	2280	2289	2298	2307	2316	2325	2334	2343	2352	2361	2370	2379	2388	2397	2406	2415	2424	2433	2442	2451	2460	2469	2478	2487	2496	2505	2514	2523	2532	2541	2550	2559	2568	2577	2586	2595	2604	2613	2622	2631	2640	2649	2658	2667	2676	2685	2694	2703	2712	2721	2730	2739	2748	2757	2766	2775	2784	2793	2802	2811	2820	2829	2838	2847	2856	2865	2874	2883	2892	2901	2910	2919	2928	2937	2946	2955	2964	2973	2982	2991	3000	3009	3018	3027	3036	3045	3054	3063	3072	3081	3090	3099	3108	3117	3126	3135	3144	3153	3162	3171	3180	3189	3198	3207	3216	3225	3234	3243	3252	3261	3270	3279	3288	3297	3306	3315	3324	3333	3342	3351	3360	3369	3378	3387	3396	3405	3414	3423	3432	3441	3450	3459	3468	3477	3486	3495	3504	3513	3522	3531	3540	3549	3558	3567	3576	3585	3594	3603	3612	3621	3630	3639	3648	3657	3666	3675	3684	3693	3702	3711	3720	3729	3738	3747	3756	3765	3774	3783	3792	3801	3810	3819	3828	3837	3846	3855	3864	3873	3882	3891	3900	3909	3918	3927	3936	3945	3954	3963	3972	3981	3990	3999	4008	4017	4026	4035	4044	4053	4062	4071	4080	4089	4098	4107	4116	4125	4134	4143	4152	4161	4170	4179	4188	4197	4206	4215	4224	4233	4242	4251	4260	4269	4278	4287	4296	4305	4314	4323	4332	4341	4350	4359	4368	4377	4386	4395	4404	4413	4422	4431	4440	4449	4458	4467	4476	4485	4494	4503	4512	4521	4530	4539	4548	4557	4566	4575	4584	4593	4602	4611	4620	4629	4638	4647	4656	4665	4674	4683	4692	4701	4710	4719	4728	4737	4746	4755	4764	4773	4782	4791	4800	4809	4818	4827	4836	4845	4854	4863	4872	4881	4890	4899	4908	4917	4926	4935	4944	4953	4962	4971	4980	4989	4998	5007	5016	5025	5034	5043	5052	5061	5070	5079	5088	5097	5106	5115	5124	5133	5142	5151	5160	5169	5178	5187	5196	5205	5214	5223	5232	5241	5250	5259	5268	5277	5286	5295	5304	5313	5322	5331	5340	5349	5358	5367	5376	5385	5394	5403	5412	5421	5430	5439	5448	5457	5466	5475	5484	5493	5502	5511	5520	5529	5538	5547	5556	5565	5574	5583	5592	5601	5610	5619	5628	5637	5646	5655	5664	5673	5682	5691	5700	5709	5718	5727	5736	5745	5754	5763	5772	5781	5790	5799	5808	5817	5826	5835	5844	5853	5862	5871	5880	5889	5898	5907	5916	5925	5934	5943	5952	5961	5970	5979	5988	5997	6006	6015	6024	6033	6042	6051	6060	6069	6078	6087	6096	6105	6114	6123	6132	6141	6150	6159	6168	6177	6186	6195	6204	6213	6222	6231	6240	6249	6258	6267	6276	6285	6294	6303	6312	6321	6330	6339	6348	6357	6366	6375	6384	6393	6402	6411	6420	6429	6438	6447	6456	6465	6474	6483	6492	6501	6510	6519	6528	6537	6546	6555	6564	6573	6582	6591	6600	6609	6618	6627	6636	6645	6654	6663	6672	6681	6690	6699	6708	6717	6726	6735	6744	6753	6762	6771	6780	6789	6798	6807	6816	6825	6834	6843	6852	6861	6870	6879	6888	6897	6906	6915	6924	6933	6942	6951	6960	6969	6978	6987	6996	7005	7014	7023	7032	7041	7050	7059	7068	7077	7086	7095	7104	7113	7122	7131	7140	7149	7158	7167	7176	7185	7194	7203	7212	7221	7230	7239	7248	7257	7266	7275	7284	7293	7302	7311	7320	7329	7338	7347	7356	7365	7374	7383	7392	7401	7410	7419	7428	7437	7446	7455	7464	7473	7482	7491	7500	7509	7518	7527	7536	7545	7554	7563	7572	7581	7590	7599	7608	7617	7626	7635	7644	7653	7662	7671	7680	7689	7698	7707	7716	7725	7734	7743	7752	7761	7770	7779	7788	7797	7806	7815	7824	7833	7842	7851	7860	7869	7878	7887	7896	7905	7914	7923	7932	7941	7950	7959	7968	7977	7986	7995	8004	8013	8022	8031	8040	8049	8058	8067	8076	8085	8094	8103	8112	8121	8130	8139	8148	8157	8166	8175	8184	8193	8202	8211	8220	8229	8238	8247	8256	8265	8274	8283	8292	8301	8310	8319	8328	8337	8346	8355	8364	8373	8382	8391	8400	8409	8418	8427	8436	8445	8454	8463	8472	8481	8490	8499	8508	8517	8526	8535	8544	8553	8562	8571	8580	8589	8598	8607	8616	8625	8634	8643	8652	8661	8670	8679	8688	8697	8706	8715	8724	8733	8742	8751	8760	8769	8778	8787	8796	8805	8814	8823	8832	8841	8850	8859	8868	8877	8886	8895	8904	8913	8922	8931	8940	8949	8958	8967	8976	8985	8994	9003	9012	9021	9030	9039	9048	9057	9066	9075	9084	9093	9102	9111	9120	9129	9138	9147	9156	9165	9174	9183	9192	9201	9210	9219	9228	9237	9246	9255	9264	9273	9282	9291	9300	9309	9318	9327	9336	9345	9354	9363	9372	9381	9390	9399	9408	9417	9426	9435	9444	9453	9462	9471	9480	9489	9498	9507	9516	9525	9534	9543	9552	9561	9570	9579	9588	9597	9606	9615	9624	9633	9642	9651	9660	9669	9678	9687	9696	9705	9714	9723	9732	9741	9750	9759	9768	9777	9786	9795	9804	9813	9822	9831	9840	9849	9858	9867	9876	9885	9894	9903	9912	9921	9930	9939	9948	9957	9966	9975	9984	9993	10002	10011	10020	10029	10038	10047	10056	10065	10074	10083
----	--------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------

FT DOMAIN 2127 EGF-LIKE 36, CALCIUM-BINDING. 2165
FT DOMAIN 2166 EGF-LIKE 37, CALCIUM-BINDING. 2205
FT DOMAIN 2206 EGF-LIKE 38, CALCIUM-BINDING. 2245
FT DOMAIN 2247 EGF-LIKE 39, CALCIUM-BINDING. 2290
FT DOMAIN 2291 EGF-LIKE 40, CALCIUM-BINDING. 2332
FT DOMAIN 2333 TGF- 7. 2400
FT DOMAIN 2402 EGF-LIKE 41, CALCIUM-BINDING. 2443
FT DOMAIN 2444 EGF-LIKE 42, CALCIUM-BINDING. 2484
FT DOMAIN 2485 EGF-LIKE 43, CALCIUM-BINDING. 2523
FT DOMAIN 2524 EGF-LIKE 44, CALCIUM-BINDING. 2566
FT DOMAIN 2567 EGF-LIKE 45, CALCIUM-BINDING. 2606
FT DOMAIN 2607 EGF-LIKE 46, CALCIUM-BINDING. 2647
FT DOMAIN 2648 EGF-LIKE 47, CALCIUM-BINDING. 2687
FT DISULFID 85 94
FT DISULFID 89 100
FT DISULFID 102 111
FT DISULFID 119 129
FT DISULFID 123 134
FT DISULFID 136 145
FT DISULFID 150 160
FT DISULFID 154 166
FT DISULFID 168 177
FT DISULFID 250 262
FT DISULFID 257 271
FT DISULFID 273 286
FT DISULFID 292 304
FT DISULFID 299 313
FT DISULFID 315 328
FT DISULFID 453 465
FT DISULFID 460 474
FT DISULFID 476 488
FT DISULFID 494 504
FT DISULFID 499 513
FT DISULFID 515 528
FT DISULFID 534 546
FT DISULFID 541 555
FT DISULFID 557 570
FT DISULFID 576 587
FT DISULFID 582 596
FT DISULFID 598 611
FT DISULFID 617 628
FT DISULFID 623 637
FT DISULFID 639 652
FT DISULFID 727 739
FT DISULFID 734 748
FT DISULFID 750 763
FT DISULFID 769 781
FT DISULFID 776 790
FT DISULFID 792 805
FT DISULFID 811 821
FT DISULFID 816 830
FT DISULFID 832 845
FT DISULFID 914 926
FT DISULFID 931 935
FT DISULFID 937 950
FT DISULFID 1032 1044
FT DISULFID 1039 1053
FT DISULFID 1055 1068
FT DISULFID 1074 1086
FT DISULFID 1081 1095
FT DISULFID 1097 1111
FT DISULFID 1117 1129
FT DISULFID 1124 1138
FT DISULFID 1140 1153
FT DISULFID 1159 1171
FT DISULFID 1166 1180
FT DISULFID 1182 1195
FT DISULFID 1201 1212
FT DISULFID 1208 1221
FT DISULFID 1223 1236
FT DISULFID 1242 1254
FT DISULFID 1249 1263
FT DISULFID 1265 1278

FT DISULFID 1284 1296 BY SIMILARITY.
FT DISULFID 1291 1305 BY SIMILARITY.
FT DISULFID 1307 1320 BY SIMILARITY.
FT DISULFID 1326 1339 BY SIMILARITY.
FT DISULFID 1333 1348 BY SIMILARITY.
FT DISULFID 1350 1361 BY SIMILARITY.
FT DISULFID 1367 1380 BY SIMILARITY.
FT DISULFID 1374 1389 BY SIMILARITY.
FT DISULFID 1391 1402 BY SIMILARITY.
FT DISULFID 1408 1420 BY SIMILARITY.
FT DISULFID 1415 1429 BY SIMILARITY.
FT DISULFID 1431 1444 BY SIMILARITY.
FT DISULFID 1450 1461 BY SIMILARITY.
FT DISULFID 1456 1470 BY SIMILARITY.
FT DISULFID 1472 1485 BY SIMILARITY.
FT DISULFID 1491 1502 BY SIMILARITY.
FT DISULFID 1497 1511 BY SIMILARITY.
FT DISULFID 1513 1526 BY SIMILARITY.
FT DISULFID 1610 1622 BY SIMILARITY.
FT DISULFID 1617 1631 BY SIMILARITY.
FT DISULFID 1633 1646 BY SIMILARITY.
FT DISULFID 1652 1663 BY SIMILARITY.
FT DISULFID 1658 1672 BY SIMILARITY.
FT DISULFID 1674 1687 BY SIMILARITY.

Query Match 3.1%; Score 173.5; DB 1; Length 2871;
Best Local Similarity 19.4%; Pred. No. 0.00025;
Matches 216; Conservative 111; Mismatches 342; Indels 443; Gaps 67;

QY 35 CCMALAGCQAAW----AGDLPSSSSRPLPPCQEKDYHFEYTECDSSGSRMVA-IPNSAV 89
Db 980 CCCSVG---AAMGTECECECLNRSR-----EVEELCPGPGFATKDTNGKP 1024

QY 90 -----DCSGLPDPVRGEC-----TFSCA--SGEYLEMKNQVCSK-----CGEG 126
Db 1025 FFKDINECKMIPSLCTHGHKCRNTIGSFKCRDSGFALDSEERNCTDIDECRISPDLCGRG 1084

QY 127 TVSLGSG---IRFDEWDELPAGFSNIATPMOTVWGPSPD---SRPDGNNSSW-----IPR 175
Db 1085 QCVNTPGDFECKDEGYE--SGFMNMKNMCDIDECORDPLLCRGGGICHNTGTYRCECPP 1142

QY 176 GNYIESN-----RDDCTVSLIYAVHLKAGSYVFEYQYVDNNIFFE-----FFQNDQC 224
Db 1143 GHQLSPNISACIDINECELSANLCPH--GRCVNLIGKYQCACNPGVHPTHDRFLFCVDIDEC 1201

QY 225 QEMDTTQKWLKTONGEWGSHSVMLKSGTNILYWRHTTCLMGSKAVKPLVKNITIEGV 284
Db 1202 SIMNGGCETFCFCTNSD-----GSY-----1219

QY 285 AYTSECFCKPGTFSNKGPSFNCQVCPNRTYSEKAKECIRCKDDSQFSGSECTERP-- 342
Db 1220 ----EC-SCQPG-FALMPDORSC-----TDIDQEDNPNCIDGGCQCTNIPGE 1260

QY 343 -PCTTKDYQIHTPCDEEKTQIMYKW--IEPKICREDLTDAIR-----LPPSGEK- 390
Db 1261 YRCLCYDGFH-----ASEDMKTCVDVNECDLNPNCISLGTCTKGSFICHDGMYGSKKG 1316

QY 391 ----KDCPPCPNPGFYN-----NGSSSCHPCPPGPFSDGTK--ECRCPGACTEPAL 434
Db 1317 KTGCTDINECELGAINCGRHVACTNTAGSFKC--SCSPGHGIGGIGKICTDLDECSNGTHMC- 1374

QY 435 GFKEYKWNVLPGNMKTSFCFN-VGNSKC---DGMNGWEVAGDHIQSGAGSDNDYLTILNIH 490
Db 1375 -----SQUADCKNTNGSYRCLCKDCYTG-----DGFCTDLDECSNIN 1413

QY 491 I-----PGKPTPTMTGATGSELGRITF-----VFET-----LCSAD 522
Db 1414 LCGNQCLNAPGGYRCECDMGFVP--SADGKACEIDECSLPNICVFCFCHNLPGLFCE 1471

QY 523 CVLYFEMVDINRKSTNVVESWGCTKEKQAYTHLIFKNATFT---FT-----WAFPTNOG 573
Db 1472 CEIGY--ELDRSGGNCID-----VNECLDPTTICSGNCVNTPGSYTCDSPDELNPTRVG 1525

```

QY 574 QNRRFINDMKYI-----SITATNAV-DGVASSCRACALG-----SEQ 611
DB 1526 CVDRSGCNVLRPRGNDGTACSGNEIGVSKASCCSILGRAGWGPCLPCPSVNTSEY 1585
QY 612 -----SGSSCV-----PCPPGHVIEKETNQ 631
DB 1586 KILCPGEGFRPNITVILEIDECQELPLGCGGKINTFGSFQCRPTGYVLNEDTRV 1645
QY 632 CKE-----CPPDYLSHQVYKREACIPCGP-----GSKNND--HSVCYDCFF 674
DB 1646 CDDVNECETPCGPGT--CYNTVGNVTCI--CPDYMGVNGNCCMDMRRSICR-----1697
QY 675 YHEKENOILHYDSNLSVSGSLMGNSPFTSKGTYKFFHFNISLCGHEGKKALCTNNITD 734
DB 1698 -----NYADNOTCDGELL-----FNNT-----KKMCCCSYNIR 1727
QY 735 FTVKEI-----VAGSDDDYTLNLVGAFCVQSTIPSESKGFRALSSQSILLADTIGVTVET 790
DB 1728 ANWKPCQCPSTDEPATLCG-----SORPGF-----VIDIYTGLPVD-1766
QY 791 TLKNINKEDMFVPTSQIP-----DVHFFYKSSSTAT-----TSCIN 827
DB 1767 -----IDE-----CREIPGVCNGVCINMGVSFRCEPCVGFYNDKLLVCEDIDECON 1814
QY 828 GR-----STAVKMRNPTKSGAGVIS-----VPSKCPACTCDGC--TFY 864
DB 1815 GPVCLRNACINTAGSYRCD-CKPGYRLTSTGQCDNRNECOEIPNICSHGQCIDTVGSFY 1873
QY 865 FL-----WESAACPLCTEHDHFHIE-GACKRG 891
DB 1874 CLCHTGKTNEDQTMCL--DINECERDAGNG 1903

RESULT 14
ID LMB1_MOUSE
AC P02469;
DC 21-JUL-1986 (Rel. 01, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
GN Laminin beta-1 chain precursor (Laminin B1 chain).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87147212; PubMed=3493487;
RA Sasaki M., Kato S., Kohno K., Martin G.R., Yamada Y.;
RT "Sequence of the cDNA encoding the laminin B1 chain reveals a
RT multidomain protein containing cysteine-rich repeats.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:935-939(1987).
RN [2]
RP SEQUENCE OF 1292-1786 FROM N.A.
RX MEDLINE=85051302; PubMed=6209134;
RA Barlow D.P., Green N.M., Kurkinen M., Hogan B.L.M.;
RT "Sequencing of laminin B chain cDNAs reveals C-terminal regions of
RT coiled-coil alpha-helix.";
RL EMBO J. 3:2355-2362(1984).
RN [3]
RP SEQUENCE OF 165-172; 539-547 AND 712-719.
RC STRAIN=BALB/c; TISSUE=Endothelial cells;
RX MEDLINE=97363207; PubMed=9219532;
RA Frieser M., Noeckel H., Pausch F., Roeder C., Hahn A., Deutzmann R.,
RA Sorokin L.M.;
RT "Cloning of the mouse laminin alpha 4 cDNA. Expression in a subset of
RT endothelium.";
RL Eur. J. Biochem. 246:727-735(1997).
CC -!- FUNCTION: Binding to cells via a high affinity receptor, laminin
CC is thought to mediate the attachment, migration, and organization
CC of cells into tissues during embryonic development by interacting
CC with other extracellular matrix components.

```

```

CC -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
CC different polypeptide chains (alpha, beta, gamma), which are bound
CC to each other by disulfide bonds into a cross-shaped molecule
CC comprising one long and three short arms with globules at each
CC end.
CC THE BETA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ), LAMININ-
CC 2 (MEROSIN), AND LAMININ-6 (K-LAMININ).
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
CC COMPONENT).
CC -!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -!- SIMILARITY: CONTAINS 12.5 LAMININ EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; M1525; AAA39407.1; ALT_INIT.
CC EMBL; X05212; CAA28839.1; -.
CC PIR; A26413; MMSB1.
CC HSSP; P02468; 1KLO.
CC MGD; MGI:96743; Lambl-1.
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR001886; LamNT.
CC InterPro; IPR002049; Laminin_EGF.
CC Pfam; PF00053; laminin_EGF; 13.
CC Pfam; PF00055; laminin_Nterm; 1.
CC PRINTS; PR00011; EGF_LAMININ.
CC PRODOM; PD002082; LamNT; 1.
CC SMART; SM00180; EGF_Lam; 11.
CC SMART; SM00136; LamNT; 1.
CC PROSITE; PS00022; EGF_1; 9.
CC PROSITE; PS01186; EGF_2; 2.
CC PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
CC Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
CC Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
CC SIGNAL 1 21
CC CHAIN 22 1786 LAMININ BETA-1 CHAIN.
CC DOMAIN 22 270 LAMININ N-TERMINAL (DOMAIN VI).
CC DOMAIN 271 334 LAMININ EGF-LIKE 1.
CC DOMAIN 335 397 LAMININ EGF-LIKE 2.
CC DOMAIN 398 457 LAMININ EGF-LIKE 3.
CC DOMAIN 458 509 LAMININ EGF-LIKE 4.
CC DOMAIN 510 540 LAMININ EGF-LIKE 5 (INCOMPLETE).
CC DOMAIN 541 772 LAMININ DOMAIN IV.
CC DOMAIN 773 820 LAMININ EGF-LIKE 6.
CC DOMAIN 821 866 LAMININ EGF-LIKE 7.
CC DOMAIN 867 916 LAMININ EGF-LIKE 8.
CC DOMAIN 917 975 LAMININ EGF-LIKE 9.
CC DOMAIN 976 1027 LAMININ EGF-LIKE 10.
CC DOMAIN 1028 1083 LAMININ EGF-LIKE 11.
CC DOMAIN 1084 1131 LAMININ EGF-LIKE 12.
CC DOMAIN 1132 1178 LAMININ EGF-LIKE 13.
CC DOMAIN 1179 1397 DOMAIN II.
CC DOMAIN 1398 1430 DOMAIN ALPHA.
CC DOMAIN 1431 1786 DOMAIN I.
CC DOMAIN 1216 1315 COILED COIL (POTENTIAL).
CC DOMAIN 1368 1388 COILED COIL (POTENTIAL).
CC DOMAIN 1448 1778 COILED COIL (POTENTIAL).
CC DOMAIN 271 280 BY SIMILARITY.
CC DOMAIN 273 298 BY SIMILARITY.
CC DOMAIN 300 309 BY SIMILARITY.
CC DOMAIN 312 332 BY SIMILARITY.
CC DOMAIN 335 344 BY SIMILARITY.
CC DOMAIN 337 362 BY SIMILARITY.
CC DOMAIN 365 374 BY SIMILARITY.
CC DOMAIN 375 395 BY SIMILARITY.
CC DOMAIN 398 411 BY SIMILARITY.

```

FT DISULFID 400 BY SIMILARITY. 426
FT DISULFID 428 BY SIMILARITY. 437
FT DISULFID 440 BY SIMILARITY. 455
FT DISULFID 458 BY SIMILARITY. 472
FT DISULFID 460 BY SIMILARITY. 479
FT DISULFID 481 BY SIMILARITY. 490
FT DISULFID 493 BY SIMILARITY. 507
FT DISULFID 773 BY SIMILARITY. 785
FT DISULFID 775 BY SIMILARITY. 792
FT DISULFID 794 BY SIMILARITY. 803
FT DISULFID 806 BY SIMILARITY. 818
FT DISULFID 821 BY SIMILARITY. 833
FT DISULFID 823 BY SIMILARITY. 840
FT DISULFID 842 BY SIMILARITY. 851
FT DISULFID 854 BY SIMILARITY. 864
FT DISULFID 867 BY SIMILARITY. 876
FT DISULFID 869 BY SIMILARITY. 883
FT DISULFID 886 BY SIMILARITY. 895
FT DISULFID 898 BY SIMILARITY. 914
FT DISULFID 919 BY SIMILARITY. 933
FT DISULFID 946 BY SIMILARITY. 944
FT DISULFID 946 BY SIMILARITY. 955
FT DISULFID 958 BY SIMILARITY. 973
FT DISULFID 976 BY SIMILARITY. 990
FT DISULFID 978 BY SIMILARITY. 997
FT DISULFID 1000 BY SIMILARITY. 1009
FT DISULFID 1012 BY SIMILARITY. 1025
FT DISULFID 1084 BY SIMILARITY. 1096
FT DISULFID 1086 BY SIMILARITY. 1103
FT DISULFID 1105 BY SIMILARITY. 1114
FT DISULFID 1117 BY SIMILARITY. 1129
FT DISULFID 1132 BY SIMILARITY. 1144
FT DISULFID 1134 BY SIMILARITY. 1151
FT DISULFID 1134 BY SIMILARITY. 1162
FT DISULFID 1165 BY SIMILARITY. 1176
FT DISULFID 1179 BY SIMILARITY. 1179
FT DISULFID 1182 BY SIMILARITY. 1182
FT DISULFID 1785 BY SIMILARITY. 1785
FT CARBOHYD 120 (POTENTIAL). 120
FT CARBOHYD 356 (POTENTIAL). 356
FT CARBOHYD 356 (POTENTIAL). 356
FT CARBOHYD 519 (POTENTIAL). 519
FT CARBOHYD 677 (POTENTIAL). 677
FT CARBOHYD 1041 (POTENTIAL). 1041
FT CARBOHYD 1195 (POTENTIAL). 1195
FT CARBOHYD 1279 (POTENTIAL). 1279
FT CARBOHYD 1336 (POTENTIAL). 1336
FT CARBOHYD 1343 (POTENTIAL). 1343
FT CARBOHYD 1487 (POTENTIAL). 1487
FT CARBOHYD 1533 (POTENTIAL). 1533
FT CARBOHYD 1542 (POTENTIAL). 1542
FT CARBOHYD 1643 (POTENTIAL). 1643
FT CONFLICT 1531 (POTENTIAL). 1531
FT CONFLICT 1749 (POTENTIAL). 1749
SQ SEQUENCE 1786 AA: 196904 MW: 846671B7BF41A474 CRC64;

Query Match 3.0%; Score 173; DB 1; Length 1786;
Best Local Similarity 18.1%; Pred. No. 0.00015;
Matches 237; Conservative 135; Mismatches 444; Indels 494; Gaps 65;

QY 61 COE-KDYHFEYTE-----CDSSGRWRVAIPNSAVDCSLPDPVRGECTFS 106
DB 374 CEQCKPEYQHPDIRDPNLCPECTCDPAGSE-----NGGI-CDGYTD-----FS 418

QY 107 CA--SGE--YLEMKNOVCSKGGTYSLSGSIKFEWDELPAFNSIA-TFMDTVVG-- 158
DB 419 VGLTAGQCRKLVHEGERCOVCKEGFDLSA-----EDPYCKSCACNPLQTIPIGPN 470

QY 159 PSDSRP-----DGC-----NNS-----S 171
DB 471 PCDSETGYCYCKRLVGTQRCDQLQPHWGLNSLDGCRPCDCLGGLNNSCEDSDSGQS 530

QY 172 WIP-----RGNVIESNRDDCTVS-LIYAVHLKK--SGVVFPEYQYVDNNI----- 213

DB 531 CLPHMIGRCQNEVESGYFTTLDHYIYEAENLPGCVVVVVERQYIODRIPSWTGPCFVR 590
QY 214 -----FEEFIQN-DOCOEMDTTDDKWKLTDMGECGSHVMLKSGTNILYWTGTLIMG 267
DB 591 VPEGAYLEFFIDNIPYSMEYEILIRVEPQLPDHWE----- 625
QY 268 SKAVKPVLVNITIE---GVAYTSECPCPKPG---TFSNKPGEFNCQVCPRTYTSKGA 320
DB 626 -KAV-----ITVQRPCKTPASSRCNGTVPDDDDNQVVSLSPGS-RYVVLPRPVCFEKGM 676
QY 321 KECIRCKDDSOFGSSECTE-----RPECCTTKDYFOIHTPCDEE-----GKTQI 364
DB 677 NYTVRL-ELQYTAGSDVESPYTFIDSLVLMPCYKSLDIFTVGGSGDGEVTSNAWETFF 735
QY 365 MYKWIE--PKICREDLTDAIR-----LPPSGEKKDCPP-----CNPGFYNNCSS 406
DB 736 RYRCLENSRSVVKTPMTDVCNIIIFSIHALIHTQCLACEDCPQSGLSVSCDP---NGQC 792
QY 407 SCHP-----CPPGTFSDGTKECRPCAPAGTEPALGFEYKMMNVLPGNMKTSCF-NVG 456
DB 793 QCRPNVVGRTCNRCAPGTGFGNGCKPCDCHLO---GSASAFCAITAIG---QCHCFQGIY 847
QY 457 NSKCD-----GMNGWEVAGDHIOGAGGSDNDYILILNLIHIFGFPKPPPTSMT 501
DB 848 ARQCDRLPGYWGFPSCQPCOCNGIHALDCTV-TGECLSCODY-----TTCHNCERCLA 900
QY 502 G-----ATGSELGR-----ITVFETLCSA-----DCVL 525
DB 901 GYVGDPIIGSGDHCRPCPCPDGPDGQFARSCYQDPVTLQACVCDPGYTGSRDDDCAS 960
QY 526 YFVMDINKSTNVVESMGTKKQAYTHIFKNATFTTFAFORTNQODNRFRINDMVK 585
DB 961 GFF-----GNPSDFGSCQPCQCHNI----- 982
QY 586 IYSITATNAVQVASSCRACALGSESGSSVPCPPGHY----- 624
DB 983 --DTTDEACDDTKRCLKCLYHTE--GDHCLQCYGYGADLRDRCRCVCNYLGTVKE 1038
QY 625 -----IEKTNCK-----ECPDPTVLSIHOVYKEACTPC-----GP 657
DB 1039 HCGSDCHDKATGQCSCLPNVIGQNCDCRCAPTW---QLASGTGCGPCNCAHAFGP 1094
QY 658 GSKNNQDHVCY-----SDC--FFYHEKENQILHYDFSNL-----SSVSLMNG 699
DB 1095 SCNEFTGOCQCPMPGGRTRCSECOELFWGDPDVECRACDPRGIETPOCDQSTGOC--- 1151
QY 700 PSFTSGTKTYFHFNFISLCGHEGCKMALCTNITDETVKEIVAGSDDYTNLVGAFVCOST 759
DB 1152 -----VCVEGVEGPRCDKCTRGYSV-----PPDCTPCQCPCFALWDA 1188
QY 760 IIPSESKGFRAALSSQSIILADTFIG---VTVTTLKLNIN-IKEDMFPVPTSO-IPDVHP 814
DB 1189 IIGELNTRTHKLEKAKALKISGVIGPYRETVDVSEKVKYNEIKDILAQSPAAPLANIGI 1248
QY 815 FYKSTATTSCINGRSTAVKMRNPT---KSGAGVIS----- 848
DB 1249 LPFEAEKLTQDVTKEKMAQVEVKLUTDTASQSNSTAGELGALQAEAEISLDTKVELAQLEF 1308
QY 849 VPSKCPAGTCDCGTFTFLW-----ESAACPLC---TEHDPHEIEGACK 889
DB 1309 IKNSDIQGALDSTTKYFQMSLEAEKRVNASTTDPNPNSTVQEOSALTDRKVEDLMLESPPK 1368
QY 890 RGQETLYVNEPKWCIFKISLPEKKLATCET-----VDFWLKV--GA 930
DB 1369 EQOEQARLUDELAKLOSLDLSAQAQMTCTGTPGADCSCECGGPNCRDTECEKCGGP 1428
QY 931 GVGAFTAVLLVALTCYFWKKNQKLE-----YKSKLVMTTNSKECELPAADSCAI 980
DB 1429 CGCG-----LVTVAHSAWQKAMDFDROVLSALAEVEQLSKMVSSEAKVRADEAKQAOVDL 1483
QY 981 MEGEDNEEEVYVSNK--QSLIGKLSLATKEKE--DHVESVQLKTSRSPN 1026
DB 1484 LKTNATREKVDKSNEDLRNLKIQIRNFLTEDSADLSDTEAVANEVLKSGN 1533

Receptor; Transcription regulation; Activator; Differentiation; Developmental protein; Neurogenesis; Repeat; ANK repeat; EGF-like domain; Transmembrane; Glycoprotein; Signal.			
SIGNAL	1	20	POTENTIAL
CHAIN	21	2437	NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1.
DOMAIN	21	1724	EXTRACELLULAR (POTENTIAL).
DOMAIN	1725	1747	POTENTIAL.
DOMAIN	1748	2437	CYTOPLASMIC (POTENTIAL).
DOMAIN	21	57	EGF-LIKE 1.
DOMAIN	58	98	EGF-LIKE 2.
DOMAIN	101	138	EGF-LIKE 3.
DOMAIN	139	175	EGF-LIKE 4.
DOMAIN	177	215	EGF-LIKE 5.
DOMAIN	217	254	EGF-LIKE 6.
DOMAIN	256	292	EGF-LIKE 7.
DOMAIN	294	332	EGF-LIKE 8.
DOMAIN	334	370	EGF-LIKE 9.
DOMAIN	371	409	EGF-LIKE 10.
DOMAIN	411	449	EGF-LIKE 11.
DOMAIN	451	487	EGF-LIKE 12.
DOMAIN	489	524	EGF-LIKE 13.
DOMAIN	526	562	EGF-LIKE 14.
DOMAIN	564	599	EGF-LIKE 15.
DOMAIN	601	637	EGF-LIKE 16.
DOMAIN	639	674	EGF-LIKE 17.
DOMAIN	676	712	EGF-LIKE 18.
DOMAIN	714	749	EGF-LIKE 19.
DOMAIN	751	787	EGF-LIKE 20.
DOMAIN	789	825	EGF-LIKE 21.
DOMAIN	827	865	EGF-LIKE 22.
DOMAIN	867	903	EGF-LIKE 23.
DOMAIN	905	941	EGF-LIKE 24.
DOMAIN	943	979	EGF-LIKE 25.
DOMAIN	981	1017	EGF-LIKE 26.
DOMAIN	1019	1055	EGF-LIKE 27.
DOMAIN	1057	1093	EGF-LIKE 28.
DOMAIN	1095	1141	EGF-LIKE 29.
DOMAIN	1143	1179	EGF-LIKE 30.
DOMAIN	1181	1217	EGF-LIKE 31.
DOMAIN	1219	1263	EGF-LIKE 32.
DOMAIN	1265	1303	EGF-LIKE 33.
DOMAIN	1305	1344	EGF-LIKE 34.
DOMAIN	1346	1382	EGF-LIKE 35.
DOMAIN	1385	1423	EGF-LIKE 36.
REPEAT	1446	1486	LIN/NOTCH 1.
REPEAT	1487	1520	LIN/NOTCH 2.
REPEAT	1521	1561	LIN/NOTCH 3.
REPEAT	1867	1910	ANK 1.
REPEAT	1915	1944	ANK 2.
REPEAT	1948	1978	ANK 3.
REPEAT	1982	2011	ANK 4.
REPEAT	2015	2044	ANK 5.
REPEAT	2048	2077	ANK 6.
DOMAIN	2265	2276	POLY-GLN (OPA-REPEAT).
SITE	1656	1657	CLEAVAGE BY FURIN-LIKE PROTEASE (BY SIMILARITY).
DISULFID	25	35	BY SIMILARITY.
DISULFID	29	45	BY SIMILARITY.
DISULFID	47	56	BY SIMILARITY.
DISULFID	62	73	BY SIMILARITY.
DISULFID	67	86	BY SIMILARITY.
DISULFID	88	97	BY SIMILARITY.
DISULFID	105	116	BY SIMILARITY.
DISULFID	110	126	BY SIMILARITY.
DISULFID	128	137	BY SIMILARITY.
DISULFID	143	154	BY SIMILARITY.
DISULFID	148	163	BY SIMILARITY.
DISULFID	165	174	BY SIMILARITY.
DISULFID	181	194	BY SIMILARITY.
DISULFID	188	203	BY SIMILARITY.
DISULFID	205	214	BY SIMILARITY.
DISULFID	221	232	BY SIMILARITY.
DISULFID	226	242	BY SIMILARITY.

FT	DISULFID	244	253	BY SIMILARITY.	Db	230	-----CTCVQTSDT-----THTCSCLPFGTGTCEHNVDDCTQHACE	266
FT	DISULFID	260	271	BY SIMILARITY.	Qy	191	-----IYAVHLKSGVVEFYQYVDNNIFFEFIQNDQCEMDT-----TTDKWVKL	237
FT	DISULFID	282	291	BY SIMILARITY.	Db	267	NGGPCIDGINTYNCHCDK-----HWTGOYCTEDV-DECELSPNACGGTCHNTIGFHCV	321
FT	DISULFID	298	311	BY SIMILARITY.	Qy	238	TONGEWG-----SHSVMLKSGTNIYWR-----TTGIL-----	265
FT	DISULFID	305	320	BY SIMILARITY.	Db	322	CVNGWTGDDCSENIDDCASAAACSHGATCHDRVASFFCECPHGRTGLLCHLDDACISNPOC	381
FT	DISULFID	322	331	BY SIMILARITY.	Qy	266	MGSKA-VKPVLVKNITIEGVAYT-----SECF-----PCK-PGTFSPKPSFNCQVC-	310
FT	DISULFID	338	349	BY SIMILARITY.	Db	382	KGSNCDTNPVSGKAICTCPPGYTGSAQNQIDECSLGANPCEHGRCLNTKGSQCKLQ	441
FT	DISULFID	343	358	BY SIMILARITY.	Qy	311	-----PRNTYSEKGAKECIR--CKDDS-----QFSG-----SSECTERPP	343
FT	DISULFID	360	369	BY SIMILARITY.	Db	442	GYEGPR--CEMDVNECKSNPCQNDATCLDQIGGFHCICMPGYEGVFCQINSDDCASQPC	498
FT	DISULFID	375	386	BY SIMILARITY.	Qy	344	CTTKDYFOI---HTPCDEEGKQIMYKWIPEKICKEDLTDAILRP-----PSGEKKDC	393
FT	DISULFID	380	397	BY SIMILARITY.	Db	499	LNGKIDKINSFHCECP-----KGFSGSLCQVDVDECASTPCKNKAKCTDGPKNYT	549
FT	DISULFID	399	408	BY SIMILARITY.	Qy	394	PPCNPGR-----YNNGSSSCHPCPPGTFSDG-----TKECRP-----	425
FT	DISULFID	415	428	BY SIMILARITY.	Db	550	CECTPGFSGIHCELDINECASS--PCHYGVCRDGVASFCTDCRPGYTGRLCETNINECLS	607
FT	DISULFID	422	437	BY SIMILARITY.	Qy	426	-----CPAGTEPALGFYKWNVLPGNMKTC--PNVGN-----	457
FT	DISULFID	439	448	BY SIMILARITY.	Db	608	QPCRNGGTCQDRENAICTCPKGI-----TGVNCEINIDDCRKKPCDY	650
FT	DISULFID	443	455	BY SIMILARITY.	Qy	458	SKC-DGMNGWEVAGDHTQSGAGGSDN-DYLLN-----LHIGFGRPPT	498
FT	DISULFID	460	475	BY SIMILARITY.	Db	651	GKCIDKINGEYCEVCEPGYSGSMCNINIDDCALNPNCHNGTCTIDGVNSFTCLCPDGRDAT	710
FT	DISULFID	477	486	BY SIMILARITY.	Qy	499	SMTGAT-----GSELGRITFEVFTLCSA-----DCVLYEMVDINRKSTNVVESGG	544
FT	DISULFID	493	503	BY SIMILARITY.	Db	711	CLSQUINESSNPCHIGSCLDQIN-SYRCVCEAGWGRNC-----DININECLSNPCVN-GG	764
FT	DISULFID	498	512	BY SIMILARITY.	Qy	545	TKEQAYTHIIFKNAFTFTWAFORTNOGONRRFINDMVKIYSITATNAVGVASSCRA	604
FT	DISULFID	514	523	BY SIMILARITY.	Db	765	TCKDMTSGYLCRCRAGFS-----GPNCOMNINECASNPCLNQSCIDDVAGFKCN	814
FT	DISULFID	530	541	BY SIMILARITY.	Qy	605	CALGSEQSGSSC---VPCPPCHYTEKETNOCKE-----CPPDTYLSIIHQVCKE	650
FT	DISULFID	535	550	BY SIMILARITY.	Db	815	CML--PYTGEVCENVLAPCSRPC--KNGGVCRESEDFQSCNCPAGWQQTCEVDINE	870
FT	DISULFID	552	561	BY SIMILARITY.	Qy	651	ACI--PCGPGSKNNODSHVCYSDCFYHEKENQILHYDFSNLSSVSGSLMNGPSTSKCTK	708
FT	DISULFID	556	568	BY SIMILARITY.	Db	871	-CVRNPCTNG-----GVC-----ENLRGGFQCRKCPGFTG----	899
FT	DISULFID	568	578	BY SIMILARITY.	Qy	709	YHFFNISLCGHEGKKMALGTNNITDFTVKEIVAGS--DDYTNLVGAFCVOSTTIIPSESK	766
FT	DISULFID	573	587	BY SIMILARITY.	Db	900	-----ALCENDIDDCENPNPCNSGGVQCDRVN--GEFVCVL-----A	933
FT	DISULFID	589	598	BY SIMILARITY.	Qy	767	GERAALSSQSIIADTFIGVTVVETTLKLNINIKEDMFPPTSQIPDVHFFYKSTATTSCI	826
FT	DISULFID	605	616	BY SIMILARITY.	Db	934	GERGERCAEDI-----DECVSAPC-----RNGGCTDCV	962
FT	DISULFID	616	625	BY SIMILARITY.	Qy	827	NGRSTAVKMRNCPKSGAGV---ISVP-----SKC-PAGTC-DGCTFYFLWESAEACPLCT	877
FT	DISULFID	627	636	BY SIMILARITY.	Db	963	NSYT-----CSCPAGFSGICEINTPDCTESSCFNGGTCDVGISF-----SCVCLPGFT	1012
FT	DISULFID	643	653	BY SIMILARITY.	Qy	878	-----EHDFFEIE-----GACKRGFQ-----ETLYVNEPKNWIKIGISL	911
FT	DISULFID	648	662	BY SIMILARITY.	Db	1013	GNVCHDNECDRSPCRNGGSCQDGYGVKTCPHGYTGLNQCQSLVRMCDSSPCKNKGSC	1072
FT	DISULFID	664	673	BY SIMILARITY.	Qy	912	PEKKLA-TCETVDFWL-----KVGAGVGAFTAVLLVALTC	945
FT	DISULFID	680	691	BY SIMILARITY.	Db	1073	WOOGASFTCCQACSGMTGIYCDVPSVSCVAVARQQGVSAVILC	1114
FT	DISULFID	685	700	BY SIMILARITY.				
FT	DISULFID	691	700	BY SIMILARITY.				
FT	DISULFID	702	711	BY SIMILARITY.				
FT	DISULFID	718	728	BY SIMILARITY.				
FT	DISULFID	723	737	BY SIMILARITY.				
FT	DISULFID	739	748	BY SIMILARITY.				
FT	DISULFID	755	766	BY SIMILARITY.				
FT	DISULFID	760	775	BY SIMILARITY.				
FT	DISULFID	777	786	BY SIMILARITY.				
FT	DISULFID	793	804	BY SIMILARITY.				
FT	DISULFID	798	813	BY SIMILARITY.				
FT	DISULFID	815	824	BY SIMILARITY.				
FT	DISULFID	831	842	BY SIMILARITY.				
FT	DISULFID	836	853	BY SIMILARITY.				
FT	DISULFID	855	864	BY SIMILARITY.				
FT	DISULFID	871	882	BY SIMILARITY.				
FT	DISULFID	876	891	BY SIMILARITY.				
FT	DISULFID	893	902	BY SIMILARITY.				
FT	DISULFID	909	920	BY SIMILARITY.				
FT	DISULFID	914	929	BY SIMILARITY.				
FT	DISULFID	931	940	BY SIMILARITY.				
FT	DISULFID	947	958	BY SIMILARITY.				
FT	DISULFID	952	967	BY SIMILARITY.				
FT	DISULFID	969	978	BY SIMILARITY.				

Query Match 3.0%; Score 170.5; DB 1; Length 2437;
Best Local Similarity 19.7%; Pred. No. 0.00034;
Matches 233; Conservative 102; Mismatches 374; Indels 473; Gaps 71;
Qy 42 COAAWAGDL-----PSSSRPLPPCQEKDYHFEYTCBDSGSRWRVAIP-----NS 87
Db 128 CQGWGKTCQLADPCASN-----PCANG-----QCSAFESHYICTCPNPHGQTCROD 177
Qy 88 AVDCSGLPDPVR-GKEC-----TFSCA-----SGEYLEMKNQVC--SKCGEGTYSLGSGI 134
Db 178 VNECAVSPSPCRNGGTCINEVGSYLRCRPPETVGPCHQRLYQCLPSPCRSG-----229
Qy 135 KFDWELPAGFSNIATFMDTVVGPDSRDCNNSWIP--RGNVIESNRDCTVSL--190

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.
OM protein - protein search, using sw model
Run on: May 12, 2003, 12:50:23 ; Search time 43.8668 Seconds
(without alignments)
4823.927 Million cell updates/sec

Title: US-10-073-333A-4
Perfect score: 5681
Sequence: 1 MLFRAGPVRGWRGAPAE.....KEKEDHFESVLKTSRSPNI 1027

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL-21.*
1: sp-archaea.*
2: sp-bacteria.*
3: sp-fungi.*
4: sp-human.*
5: sp-invertebrate.*
6: sp-mammal.*
7: sp-mhc.*
8: sp-organelle.*
9: sp-phage.*
10: sp-plant.*
11: sp-rodent.*
12: sp-virus.*
13: sp-vertebrate.*
14: sp-unclassified.*
15: sp-rvrius.*
16: sp-bacteriap.*
17: sp-archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2624	46.2	493	4 Q96DP2	Q96dp2 homo sapien
2	1753	30.9	580	4 Q9P2M2	Q9p2m2 homo sapien
3	1305	23.0	1019	5 Q9NA40	Q9na40 caenorhabdi
4	888.5	15.6	300	11 Q8R215	Q8r215 mus musculu
5	225.5	4.0	1074	5 Q964D1	Q964d1 entamoeba h
6	213.5	3.8	1101	5 Q964D2	Q964d2 entamoeba h
7	207.5	3.7	704	5 Q9U048	Q9u048 giardia lam
8	203	3.6	709	5 Q9XTJ7	Q9xtj7 giardia lam
9	201	3.5	709	5 Q97444	Q97444 giardia lam
10	199.5	3.5	667	5 Q95WU1	Q95wul giardia lam
11	199.5	3.5	2759	5 Q45614	Q45614 caenorhabdi
12	199.5	3.5	3102	5 Q9TZR4	Q9tizr4 caenorhabdi
13	197.5	3.5	3087	5 Q8WRD2	Q8wrdr2 plasmodium
14	193.5	3.4	5374	11 Q99ND0	Q99nd0 mus musculu
15	192.5	3.4	719	5 Q9U021	Q9u021 giardia lam
16	190	3.3	769	5 Q24971	Q24971 giardia lam

17	188.5	3.3	999	4 Q9NQ36	Q9nq36 homo sapien
18	187.5	3.3	1557	5 Q8WPK9	Q8wpk9 oikopleura
19	187	3.3	1723	5 Q8WRD0	Q8wrdr0 plasmodium
20	186	3.3	1007	13 Q90ZN3	Q90zn3 gallus gall
21	185.5	3.3	719	5 Q9U019	Q9u019 giardia lam
22	184.5	3.2	2189	5 Q9BI05	Q9bi05 eimeria ten
23	184.5	3.2	2534	5 Q8WRD1	Q8wrdr1 plasmodium
24	183	3.2	1372	5 P91526	P91526 caenorhabdi
25	182	3.2	2331	5 Q9W2C6	Q9w2c6 drosophila
26	181.5	3.2	1551	5 Q9NGV4	Q9ngv4 drosophila
27	181.5	3.2	3396	5 Q9VM55	Q9vm55 drosophila
28	180.5	3.2	1289	5 Q8SSS3	Q8sss3 dictyosteli
29	179.5	3.2	718	5 Q9BI07	Q9bi07 entamoeba h
30	179.5	3.2	3567	11 Q9ES77	Q9es77 mus musculu
31	178.5	3.1	3704	5 P91904	P91904 caenorhabdi
32	175	3.1	2906	11 Q9WUH9	Q9wuh9 rattus norv
33	174.5	3.1	548	5 Q9GQ45	Q9gq45 giardia lam
34	174.5	3.1	597	5 Q07317	Q07317 giardia int
35	174.5	3.1	1316	4 Q96JU7	Q96ju7 homo sapien
36	173.5	3.1	468	5 Q9BMH3	Q9bmh3 ichthyophth
37	173	3.0	667	5 Q9ATK3	Q9atk3 giardia lam
38	173	3.0	1045	5 Q8T3A6	Q8t3a6 caenorhabdi
39	173	3.0	1070	5 Q8T3A7	Q8t3a7 caenorhabdi
40	173	3.0	1111	5 Q9XWD6	Q9xwd6 caenorhabdi
41	172.5	3.0	3857	11 Q88840	Q88840 mus musculu
42	171.5	3.0	2809	4 Q96JP8	Q96jp8 homo sapien
43	169.5	3.0	739	5 Q9GS24	Q9gs24 giardia lam
44	168	3.0	594	5 Q24970	Q24970 giardia lam
45	168	3.0	1664	5 Q9TVQ2	Q9tvq2 caenorhabdi

ALIGNMENTS

RESULT 1

Q96DP2 PRELIMINARY; PRT: 493 AA.
AC Q96DP2;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE CDNA FLJ31340 fis, clone MESAN100035, weakly similar to major surface-labeled trophozoite antigen precursor.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isoqai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK055902; BAB71041.1; -;
DR InterPro; IPR001881; EGF_Ca.
DR SMART; SM00179; EGF_CA; 1.
SQ SEQUENCE 493 AA; 54595 MW; 0AEBB2C19709B211 CRC64;

Query Match 46.2%; Score 2624; DB 4; Length 493;
Best Local Similarity 97.3%; Pred. No. 4.6e-215;
Matches 470; Conservative 4; Mismatches 7; Indels 2; Gaps 1;

QY 244 GSHSVMLKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSECPCPKGCTGTSNKPQ 303

DB 4 GNYQVLMKSGTNILYWRRTTGILMGSKAVKPVLVKNITIEGVAYTSECPCPKGCTGTSNKPQ 63

QY 304 SFNCQVCPNRTYSEKGAKECIRCKDSDQSFS--GSSECTERPPCTTKDYFQIHTPCDEEGK 361

```

Db 64 SFNCQVCPNTYSEKGAKECIRCKDDSQFSEGSSECTERPPCTTKDYFQIHTPCDEEGK 123
Qy 362 TOIMYKWIETPKICREDLTDATRLPPSEKDKDCPCNPGFYNGSSSCHPCPGTFSOGTK 421
Db 124 TOIMYKWIETPKICREDLTDATRLPPSEKDKDCPCNPGFYNGSSSCHPCPGTFSOGTK 183
Qy 422 ECRPCPACTEPALGFYKWNVLPNMMKTSFCFVNGSKDCGMNGWEVAGDHIQSGAGSD 481
Db 184 ECRPCPACTEPALGFYKWNVLPNMMKTSFCFVNGSKDCGMNGWEVAGDHIQSGAGSD 243
Qy 482 NDYLILNLHIHIFGKPPKPTSMTGATSELGRITFVFETLCSADCVLVFWVDINRKSTNVYES 541
Db 244 NDYLILNLHIHIFGKPPKPTSMTGATSELGRITFVFETLCSADCVLVFWVDINRKSTNVYES 303
Qy 542 WGGTKEKQAYTHIIFKNATFTTFAFORNODNRRFINDMKIYSITATNAVDCVASS 601
Db 304 WGGTKEKQAYTHIIFKNATFTTFAFORNODNRRFINDMKIYSITATNAVDCVASS 363
Qy 602 CRACALGSEQSSGSCVPCPPGHIYEKETNOCCEKPPDTYLSIHQVYGEACIPCGPGSKN 661
Db 364 CRACALGSEQSSGSCVPCPPGHIYEKETNOCCEKPPDTYLSIHQVYGEACIPCGPGSKN 423
Qy 662 NODHSVCYSDCFFYHEKENQIHLHYDFSLSVGLMNGSPSTSKGTYFHEFNISLCGHE 721
Db 424 NODHSVCYSDCFFYHEKENQIHLHYDFSLSVGLMNGSPSTSKGTYFHEFNISVGMR 483
Qy 722 GKK 724
Db 484 GRR 486

RESULT 2
Q9P2M2
ID Q9P2M2 PRELIMINARY; PRT: 580 AA.
AC Q9P2M2
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE KIAA1324 protein (Fragment).
GN KIAA1324
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=20181126; PubMed=10718198;
RA Nagase T., Kikuno R., Ishikawa K., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.XVI.
RT The complete sequences of 150 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res 7:65-73(2000).
DR EMBL; AB037745; BAA92562.1; -.
FT NON_TER
SQ SEQUENCE 580 AA; 63365 MW; 5E63C19265EC8E5C CRC64;

Query Match 30.9%; Score 1753; DB 4; Length 580;
Best Local Similarity 53.8%; Pred. No. 1.1e-140;
Matches 307; Conservative 103; Mismatches 157; Indels 4; Gaps 2;

Qy 440 WNNVLPNMMKTSFCFVNGSKDCGMNGWEVAGDHIQSGAGSDNDYLILNLHIHIFGKPPKPT 499
Db 1 WNNVLPNMMKTSFCFVNGSKDCGMNGWEVAGDHIQSGAGSDNDYLILNLHIHIFGKPPKPT 60
Qy 500 -MTGATSELGRITFVFETLCSADCVLVFWVDINRKSTNVYESWGGTKEKQAYTHIIFKN 558
Db 61 VMADTENKEVARITFVFETLCSVNCEDLYFMVGVNSRNTFVETWKGSKGQSTYIIIEEN 120
Qy 559 ATFTTFAFORNODNRRFINDMKIYSITATNAVDCVASSCRACALGSEQSSGSCV 618
Db 121 TTTTFAFORNODNRRFINDMKIYSITATNAVDCVASSCRACALGSEQSSGSCV 180
```

```

Qy 619 CPPGHYTEKTNOCCEKPPDTYLSIHQVYGEACIPCGPGSKNODHSVCYSDCFFYHEK 678
Db 181 CPAGYVIDRDSGTCSCPPNTILKAHPYGVACVPCPGTKNNKIHSLCTNDCCTFSRNT 240
Qy 679 ENQILHVDFSNLSSVGLMNGSPSTSKGTYFHEFNISLCGHEGKMAKALCTNNITDFTVK 738
Db 241 PTRTFNFSALANTVTLAGGPFSTSKGLKYFHIHFTLSLGGNKGKMSVCTDNVTDLRIP 300
Qy 739 EIVAGSDDTYTLNVLGAFVQCSSTIIPSESGFRAALSSOSIILADTFIGVTVEVTLKNINIK 798
Db 301 E--GESGFSKSIATVYQAVLIPEVNTGYKAGVSSQVSLADRLIGVTTDMTLDGITSP 357
Qy 799 EDMFPVPTSQIPDVHIFKYKSTATTSCINGRSTAVKMKCNPTKSGAGVISVPSKCAPCTC 858
Db 358 AELFHLESGLGIPDVHIFKYKSTATTSCINGRSTAVKMKCNPTKSGAGVISVPSKCAPCTC 417
Qy 859 DGCFTYFLWESAECPLCTEHDHFEIEGACKRGFOETLYVWNEPKWCIKIGSLPEKKLAT 918
Db 418 DGCNHFHLEWESAECPLCTEHDHFEIEGACKRGFOETLYVWNEPKWCIKIGSLPEKKLAT 477
Qy 919 CETVDLWLVGAGVGAFTAVLLVALTCYFWKKNQKLEYKSKLYMTTNSKECELPAADSC 978
Db 478 CKTIDFWLVGAGVGAFTAVLLVALTCYFWKKNQKLEYKSKLYMTTNSKECELPAADSC 537
Qy 979 AIMEGEDNEEVVYVSNKOSLLGKSLATKE 1009
Db 538 AIMEGEDNEEVVYVSNKOSLLGKSLATKE 568

RESULT 3
Q9NA40
ID Q9NA40 PRELIMINARY; PRT: 1019 AA.
AC Q9NA40
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE Y73F8A.5 protein.
GN Y73F8A.5
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Matthews L.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; AL132862; CAB70224.1; -.
DR InterPro; IPR002049; Laminin_EGF.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; UNKNOWN_1.
SQ SEQUENCE 1019 AA; 112378 MW; 6EB4094722B707E7 CRC64;

Query Match 23.0%; Score 1305; DB 5; Length 1019;
Best Local Similarity 31.4%; Pred. No. 4.1e-102;
Matches 316; Conservative 175; Mismatches 398; Indels 118; Gaps 40;

Qy 61 CQEKDYHFEETCDSSGSRWRVAIPNSAVDCSLPDPVRGKECTFSCASGEYLEMKNQVC 120
Db 25 CTDDYEFYTCNDEGERWRVAVPRGGMOCNLPTRRGLNCSFCEPCHYLDLSQHC 84
Qy 121 SKCGEGYTLGSGTKFDEWDELPAFNSIATFMDTVVPGSDSRPDG-----C-N 168
Db 85 RCPNCPGFFSLGGGIRVEEFTVLPSEFS-----VDNMDSNPDAQFSNRQSQVVECPK 135
Qy 169 NSSWIPKGNVIESNRDDCTVSLIYAVHLKSGYVFEYQYVDNN--IFFEFTFNDQCQE 226
Db 169 NSSWIPKGNVIESNRDDCTVSLIYAVHLKSGYVFEYQYVDNN--IFFEFTFNDQCQE 226
```

Db 136 EAGWVVKDGLIYIPTPCVSKLSFSAVLVRPGSEVFTYRMPNRNRLASMQVDIRNEQOCS 195
QY 227 M-DTITDKWVKLT-----DNGEWSHSVYMLKSGTNNILYWRITGILMGSRAVXPVLVK 277
Db 196 YNDVAKSMFLKTKYKKEDEEKNKGMWRKRRIELKSGANVISWIONNMGYQASNOPHIID 255
QY 278 NITIEGVAYTSCFPCKPGTFSKNGSCQVCPNRYSEKGAKEICRCKDDSQFSG--S 335
Db 256 RIDVLGLAFTRCTACPPCT--SSPGSAECIPCSGFSKSGSGGGR-PEQVSGGFS 313
QY 336 SICTERPPCTTKDYQIHTPCDEEKGKTMKWIPIKICREDLTDAILRPPSGEKDKCPP 395
Db 314 EXCIDRPPCRVDYYPVREPC--TNGSSRAVYKVLPLSICRDLPLSATKLPPPTPKTCRK 372
QY 396 CNPGFYNNSSCHPCPCPTESDGTKECPKPCAGTEPALGFYKWNVLPGNWKTSCEFNV 455
Db 373 CNPGMEKNLGVCECKKDYSDG--NSCSRCPVDIVPNVGLQYQNEVLLPKLSRCEYI 431
QY 456 GN---SKDGMNGWEVAGDHIOGAG---GSDNDYLIILMLHIPGF-----KPPTSMTGA 503
Db 432 SEDVATACNIGDAWIPSGDSLISAPLSLELGIAFE--LILSID--EGWNPPLAPKPSKTMK-- 487
QY 504 TQSELGRITFVETLCSAD--CVLYFWMDIN-----RKS--TNVFSWGTGKEKQAYTHII 555
Db 488 --VPAQVITIVETSCADESCALYIEDMSAGIKQRESFYHFLAAFNGNSNKRVMSTV 545
QY 556 FKNATFTTFWAFORT--NOGQDNRFRINDMKIYSTATN-----AVDGVASSCRACALG 608
Db 546 TANTARFWAFLRGVSGDDK---ISDETRIYAINVTNVGHRGQGGGASQCUTCP-- 600
QY 609 SQGSSCVPCPPGHIIEKETNOCKEPPDTYLSI--HQVYKGAECIPGCGSKNNQDHSV 667
Db 601 HPAGGETCVPCPAGNYMHEVTKLCVSCPLNTIINTVSSRGVSKVPCQGLTSDNGVSC 660
QY 668 CYSDCFFVHE-----KENQILHYDES-----NLSSVGLMNGSPSTSKGTYKHYEFHNI 715
Db 661 TANGKIQLNQGGTGGKNGDSSSEFYDFSPVGRSWNISGV-----RVFSREGAAYHYHFSV 714
QY 716 SLGHEGKMACTNNITDFTVKEIVAGSDDTNLVGFVQCOSTIIPSEKSGKFRALSSQ 775
Db 715 SLFPPNIK---CQGFQDNFDMIGIL--DQDKENVEG--LACRTALPSPSSNRSTAYTV 767
QY 776 SIILADTIGTVETTLNINIKEDMFVP-----TSQIPDVHFFYKS--STATTSINGRS 830
Db 768 PLLLAGRLDSISLNRKHGNSLSDEVLYSDHNDTSHPLDVFFWFPEPVSTISPACPNQ 827
QY 831 TAVKRCNPTRKSGAGVISVPSKCPAGTCGCTFFFLWESAECPLCTEHDHFEIEGACKR 890
Db 828 LVVVARCVPTKQM--EMRLPHNCPDGTGCLFVIMETAQAQCPVCSNDYETINGECLN 886
QY 891 GFOETLYVWNEPKWCIKIGSLPEKKLATCETVDFWLKVGAGVGAFTAVLL--VALTCYFWK 949
Db 887 G-KQIHSIPKHKCVITGAOSKEVA--CSAFTAFQRAILSLVLUSMIFLSIGFVC--ICR 943
QY 950 KNKLEYKSKLVMTNKECELPAADSCAIMEGBDNEE---EVVYS 993
Db 944 RNRLEYKYTRLI---ESHTGELPAVETCGLDEDEDDDELQDRVIFS 987
RESULT 4
Q8R215 PRELIMINARY; PRT: 300 AA.
AC Q8R215;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
OS Hypothetical 32.6 kDa protein (Fragment).
OC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RP [1]
SEQUENCE FROM N.A.

RC TISSUE=COLON;
RA Strausberg R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC022655; AAH22655.1;
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 300 AA; 32642 MW; 6455109C054B6CB4 CRC64;
Query Match 15.6%; Score 888.5; DB 11; Length 300;
Best Local Similarity 53.6%; Pred. No. 2.3e-67;
Matches 158; Conservative 52; Mismatches 82; Indels 3; Gaps 1;
QY 728 CTTNNTDFTVKEIVAGSDDTNLVGFVQCOSTIIPSEKSGKFRALSSQIILADTIGTV 787
Db 1 CVDNVTDLRIPDEAG---FSKSVTAYVQCVIIPSEVMGYKAGVSQPVSLADRLVGVS 57
QY 788 VETTLNINIKEDMFVPPTSQIPDVHFFYKSTATTSCINGRSTAVKMRCPNTPKSGAGVI 847
Db 58 TDMTLEGIVSPVELFHPETSGIPDIVFFFRSNDVTQSCSSGRSTTIRLCNPNKAAAGTIL 117
QY 848 SVPSKCPAGTCGCTFFFLWESAECPLCTEHDHFEIEGACKRGTETLYVWNEPKWCIK 907
Db 118 RLPSMCSGTCGCTGCGNFHFLWESAECPLCSASDYHTFVSSVAGIQKTTYMREPKLCSG 177
QY 908 GISLPEKKLATCETVDFWLKVGAGVGAFTAVLLVALTCYFWKKNQKLEYKSKLVMTNS 967
Db 178 GISLPEQRVTICKTIDFWLKGISAGTCTAILTLVTCYFWKKNQKLEYKSKLVMTNATL 237
QY 968 KCELPADSCAIMEGBDNEEVEVYNSKQSLGLKLSLATKEKEDHFEVSQVLT 1022
Db 238 KCDLPAADSCAIMEGBDVEDDEFTSKSLGKIKSFTSKRTPDGDFSVPLKTS 292
RESULT 5
Q964D1 PRELIMINARY; PRT: 1074 AA.
AC Q964D1;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Gal/GalNac lectin Ig12 (Fragment).
GN IGL2.
OS Entamoeba histolytica.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=5759;
RP [1]
SEQUENCE FROM N.A.
RC STRAIN=HMI:IMSS;
RX MEDLINE=21391855; PubMed=11500468;
RA Cheng X.J., Hughes M.A., Huston C.D., Loftus B., Gilchrist C.A.,
RA Lockhart L.A., Ghosh S., Miller-Sims V., Mann B.J., Petri W.A. Jr.,
RA Tachibana H.;
RT "Intermediate Subunit of the Gal/GalNac Lectin of Entamoeba
RT histolytica Is a Member of a Gene Family Containing Multiple CXXC
RT Sequence Motifs";
RL Infect. Immun. 69:5892-5898(2001).
DR EMBL; AF37951; AAK92362.1;
DR InterPro; IPR000561; EGF-like.
DR PROSITE; PS01186; EGF_2; UNKNOWN_1.
KW Lectin.
FT NON_TER 1
FT NON_TER 1074 1074
SQ SEQUENCE 1074 AA; 116782 MW; 1552E2D714EB450F CRC64;
Query Match 4.0%; Score 225.5; DB 5; Length 1074;
Best Local Similarity 19.5%; Pred. No. 5.1e-10;
Matches 203; Conservative 117; Mismatches 392; Indels 329; Gaps 54;
QY 107 CASGYLEKNNQVCSKCGEGYSLGSGIKFDEWDELPAQFSNIATFMTFVVGPSDRPDG 166
Db 13 CAS-----VSGACACDEG-YELKT-----ESGSGSTOKTKTKEETCKSAFSYVDG 58
QY 167 CNSSWIPRGNIESNRDDCTVSLIYAVHLKSGVYFVEYQVVDNIFEFFI--QNDQC 224
SEQUENCE FROM N.A.

```
Db 59 SDSNS--PKVCYENGKESDTSFSSNNKCKCKNG-----VDT---CESCLSKDNKDC 105
Qy 225 QE----MDTTDKVKTLDNCEWGS-----VLMKSGTNILYWRITGI--LMGSKAVKP 273
Db 106 GECVIGSTTTGKGLCDTIVTDEHAENCVGLTAKSSSKQCKCFGMYLSQSGQCTK- 164
Qy 274 VLVKNITIEG--VAYTSCFPCKPOTFSNKPFCNQVCPRTYSEKGAKEICIRKDDSQ 331
Db 165 ---KNEXIEKILQVSSCNQADGYYINT--EKCKTYPDHC--SKNSDKCNGCMGEGY 218
Qy 332 FSGSECTERPPCT---TKDYFQIHTPCDEBECKTQIMYKWIPEKICREDLT--DAI----- 382
Db 219 LNG-----TECKVCTIDNSKOL-----SEGNOCSIYNAEHCSCNKRCTVSDGVCVK 266
Qy 383 --RLPPSEKCKDCPCNPGFYNGSSSCHP-----CPPGTF--SDGTKE--CR 424
Db 267 HCRLESPTEENKNCIKDDGYFLTAGKCSPLNDGFKTSATEQKGYYLEKDGDKRCS 326
Qy 425 PCPAGTEPALGFYKWNVLPGNMK-----TSCFNVGNSK 459
Db 327 LCPDPFTECLTSQ----TPVPGKLNLSAHLTSTDGPKLPGLCLCSDDDTICY-----K 377
Qy 460 CDGMNGEVAUGH-----IQSAGGSDNDYLILNL-----HIPGFKPPTSMGTGCS 506
Db 378 CE--NGLTLNGTHCYNFDVKVLTSGNHNQVCKMRGYDQYEQYLNAFKASDN----- 428
Qy 507 ELGRITTFVEETLCS--ADCVLFYFMVDINKSTNVVSWGCTKEQOAYTHIIFKNATFTFTW 565
Db 429 -----TY-----YCPKLKDLPLPYESVTHGSDNKTIGICVGRDROV----- 464
Qy 566 AFQRTNOCODNRFT-----NDMKYISITATNAVGVASS--CRACALGSEOSGSSC 616
Db 465 ----KNDCECDKYPKSVDRKASDCVSIK--TKLPSCERAANENICTQCPVGSVDNSGK 518
Qy 617 VPCPGHYIEKTNCKECPDP--TYLSITHQVYGRACIPC-----GPKSKNQDHISVCYS 670
Db 519 CSCGDAHFD--ONNKQCBPCASCSYSSKSVVCSECYENIQGVSTRDKDNECACK 577
Qy 671 DCFYHEKENQILHYDFSNLSSVGLMNGPSPSKGTGYKRYHFNISLGCHEG----- 722
Db 578 DTPEYKEGLNA-----EDKKKSCAQLNNN-----CKEEGHYKISDGF 614
Qy 723 -----KMAICTNNITDFTVKEIVAGSDDYTNLVGAFVQSTIIPSES 765
Db 615 ITCLECDSDSAYIVDSQTECAQACSN-----AFKDNKK-----COLCSTKKDK 658
Qy 766 KGFRAALSQS--IILADTFIGVTVETTLKINIKEDMFPVPTSQIPDVHFFYKSTATTS 824
Db 659 YGHCASACATACIICEDTNVLAAGSNAOCTVCKDGFQIESPTDGY-----CSPCPAK 714
Qy 825 CINGRSTAVK--MRCNPTKSGAGVISVPSKCPA-----GTCDGCT----- 862
Db 715 KTKYKSAADKKEICVCTCQDQSVDIKPPCTACLTGTVOLENGTCQSCSDLSKYPCCKTT 774
Qy 863 -----FYFLHESAE-----CPLCTEHDHEHEGACKRGFOETLYVWNEPK 903
Db 775 DFCNVDSRTGYIATYECSDGFSGRSPYNSCTTCIESNYPKBG-----EKN 820
Qy 904 WCIGISLPEKKLATCETVDFWLKVGAGVGAFTVALLVALTCYFWKKNOKEYKYSKLV 963
Db 821 GCAK----CDDKCATCSKDKTCL-----TCTDPLKIGSKDECKTGYNSNGECK-PC 868
Qy 964 TTNSKECELPAAADSCAIMEGE 984
Db 869 TNHCSECS--SAAECTVCESD 887
```

RESULT 6
Q964D2
ID Q964D2 PRELIMINARY; PRT: 1101 AA.
AC Q964D2;
DT 01-DEC-2001 (TReMBLrel. 19, Created)

```
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)  
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
DE Gal/GalNac lectin Igl1.  
GN Entamoeba histolytica.  
OS Eukaryota; Entamoebidae; Entamoeba.  
OC NCBI_TaxID=5759;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HMI:IMSS;  
RX MEDLINE=21391855; PubMed=11500468;  
RA Cheng X.J., Hughes M.A., Huston C.D., Loftus B., Gilchrist C.A.,  
RA Lockhart L.A., Ghosh S., Miller-Sims V., Mann B.J., Petri W.A. Jr.,  
RA Tachibana H.;  
RT "Intermediate Subunit of the Gal/GalNac Lectin of Entamoeba  
RT histolytica Is a Member of a Gene Family Containing Multiple CXXC  
RL Infect. Immun. 69:5892-5898(2001).  
DR EMBL; AF377950; AAK92361.1.;  
DR InterPro; IPR000561; EGF-like.  
DR PROSITE; PS01186; EGF_2; UNKNOWN_1.  
KW Lectin.  
SQ SEQUENCE 1101 AA; 119512 MW; C8B6F5CBDE5656AEC CRC64;  
  
Query Match 3.8%; Score 213.5; DB 5; Length 1101;  
Best Local Similarity 19.9%; Pred. No. 5.6e-09;  
Matches 212; Conservative 109; Mismatches 374; Indels 371; Gaps 63;  
  
Qy 101 KECTFSCASGEVLEMKNOVCKGEGTYSLGSGIKFDEWDELPAFGSNIAFMFTVVGPS 160  
Db 27 REAVPHCAS-----VSGACATSCDTG-YELT-----TGNNKTCTLKEDMCKTA 69  
Qy 161 DSRPDGCGNNSWIPRGNYI-----ESNRDDCTVSLIYAVHLKSGVYFFQYVD 210  
Db 70 FSYVDKTNSTN--PKCTYCVNGKEVNTSSHSNDKCVCK----- 106  
Qy 211 NNIFE---EFFIIONDOCF-----MDTTDDKWKVLTNDGENGSHSV-----MLKSGTNI--- 256  
Db 107 NNVNICESCLMKSKGCEIGMSTTVYD-GSKLCDNATTEHDHAENCVGLASSTSSKTC 165  
Qy 257 -----LYWRTTGILMGSKAVKPVLVKNITIEG--VAYTSECPCKPGTFSNKPFGNFCQV 309  
Db 166 DKCFQMY-----SLOGGKCTQ-----KNDKINKILQVENSNCQCADG-YSLSTDKKSCNK 215  
Qy 310 CPRNTYSEKGAKEICIRKDDSQFSGS-SECTERPPCTTKDYFIHTPCD-BEGKTIQIMYK 367  
Db 216 FPEHC-SKINGNQCLTMEGYLSKTSKCT---ICT-----VDNPNNLSEGNCSYIN 265  
Qy 368 WLEPKICREDLT--DAI-----RLPPSGEKKDCPCPNPGFYNGSSSCHP----- 410  
Db 266 ABHCTSCNKRCTVSDGVCVKNHICRLFSPTEENKCTKCDNGYFLTTSGTCSPLNYDGFKA 325  
Qy 411 ----CPPGTF--SDGTKE--CRPCPAGTEPALGFYKWNVLPGNMK----- 449  
Db 326 NRTECENGYLEKDGDKRCSLCPDPFTECLTSK----TPVPGKLNLSAHLTSTDGPK 381  
Qy 450 -----TSCFNVGNSKCDGMNGEVAUGH-----IQSAGGSDNDYLILNL----- 489  
Db 382 LPGCLLCSDDDTICY-----KCE--NGLTLNGTHCYNFDTKSVLGTSGNHNQVCKMRGYD 434  
Qy 490 ---HIPGFKPPTSMGTGATGSELGRITTFVFE"LCASADCVL-----YFMVDINKSTNV 538  
Db 435 QYEQYLNAFKASDN-----TYY-----CPLKDLPLPYESVT----- 466  
Qy 539 VESWGTKEKQAYTHIIFKNATFTFTWAFQRTNOCQD---NRRFI-----NDMKYIS 588  
Db 467 ---KGTSN-----FITIGCVQLRNVSNDCECNKHIPTSIDKASDCVSIIT 511  
Qy 589 ITATNAVGVASSCRACALGSEOSGSSCVPCPPGHIYIEKETNOCKECPDP--TYLSIHQVY 647  
Db 512 KLPSCERTANGNICTQCPVGSIVGDKGKSCGDAHYFDKAD-NVCKKCPASCSCSYSSK 570  
Qy 648 GREACIPC-----GPKSKNQDHISVCYSDFYHEKENQILHYDFSNLSSVGLMNGPSF 702
```

```
Db 571 SKVYSECYENIOGVTTNRKNECANDGY-----KE-----GNA 607
QY 703 TSKGTKYFHFNFISILCHGEGKMACTNNITDFVKEIVAGSDDYTNLVGAFV-----CQ 757
Db 608 EDKKSQAQLNN--NCKEKG-----YEISDGV--TCLDCDSAYIVGQVACTQCS 657
QY 758 STIIPSEKGFRAALSQSIIADTFIGVTVETTLKINIKEDMFPVPTSIQIDV---HF 814
Db 658 PNAFKDENNKQCLCSTRQSOYGHCAACSATACITCEDINL-----ILTGEKPCVCKDG 711
QY 815 FYKSTAT-----TSC--INGRTAVKMR--NPTKSGAGVISVPS--KCPA-----855
Db 712 FYQIENATDGVYCPKPAKCKTKYNTTSKKVECVCTEORLKDKIAPKACPTGTQVLE 771
QY 856 -GTCDCGT-----FYFLWESAEA-----CPLCTEHDHFEI 884
Db 772 NGTCSCSLSKYPCCKKTDSCNVDSRTGFIYATECSDGSGRSPYNCNCTTKSNYYPK 831
QY 885 EGACKRGFOETLYVNMFPKWCIGISLPEKIKLATCETVDFW-----LKVGAGVGAFTAV 938
Db 832 EG-----EKNGCAK-----CDDKCATCSKDCTCLTCADPLKVGSKCDG-----869
QY 939 LVALTCYFWMKKKQLEKYKSKLWMTNSKECELPAAADSCAIMEGE 984
Db 870 ---CKTGYMSNG---ECK-----PCTNHCSECS--SAAECTVCESD 903

RESULT 7
Q9U048 PRELIMINARY; PRT; 704 AA.
ID Q9U048
AC Q9U048;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Variant-specific surface protein.
GN VSP417-6.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiae.
OX NCBI_TaxID=5741;
RN [1]
RC STRAIN=AD-1.
RA Ey P.L., Darby J.M.;
RT "A new member of the vsp417 subfamily of variant-specific surface
RL protein (VSP) genes in Giardia intestinalis."
DR EMBL: AF065606; AAF02907.1;
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR005127; Giardia_VSP.
DR Pfam: PF03302; VSP; 2.
DR SMART; SM00261; FU; 4.
SQ SEQUENCE 704 AA; 71677 MW; 7E5AE1245AD4FD45 CRC64;

Query Match 3.7%; Score 207.5; DB 5; Length 704;
Best Local Similarity 20.8%; Pred. No. 9.5e-09;
Matches 199; Conservative 88; Mismatches 353; Indels 315; Gaps 54;

QY 38 ALAGCAAWAGDLPSSSRPLPPCQEKDYHFEY--TECDSSGRWRVAIPNSAVDCSLPD 96
Db 15 ARAACQADGSGDSCKTCGVTCQE-----QVCSECGA-----NYA-----52
QY 97 PVRGKECTFSCAGEYLEM---KNQ--VCSKCG-----EGTYSLGSIGKIFDEWDELPA 145
Db 53 PVANG--QCADVNAEGQSKTCLCPAKGQKTCQCGGASFMYKDCGYS-----95
QY 146 FSNATFMDTVGSPS---DSRPGCCNNSWIPRGNYESNRDCTVSLIYAVHLKSGYV 202
Db 96 -----KDTAPGQSCMTQASEGKCTEAAPGYFLNPLRANTKDSVWSCDVTGFTDSGKT 148
QY 203 FFEYQYVDNNIFFEFTFQNDQCOEMDFTTDTKWKLLTDNGEGSHVMLKSGTNILYWR 262
Db 149 YRGVQH-----CERCQGA-----LTDAA---GGDAKTRCGQD--KYLATT 185
```

```
QY 263 GILMSKAVKPVLVKNITIEGVAYTSECPCKPGTFSNKPGRSFN---CQVCPNRYSEKG 319
Db 186 GTCG-----EG-----CTPDTEFSKEDSDNGKRCFACGDVT---TG 218
QY 320 AKECIRKDDSQFSSSECTERRPPCTTKDYFIHTPCDEEGKQIIMYKWIPEKICREDLT 379
Db 219 VASCEKCTPPSPDQAKPACTK---CGGNNY--LKTAAD--CTTTC---VQSACSPP--S 266
QY 380 DAIRLPSGGEKKDCPPONPGFYNNNGSSSCHPCPPGTSDGTGKECRPCPAGTEPALGFYK 439
Db 267 FVENSOQSGNR--CVLCGDA--ANGVDKCAACTP-----TDQGRIAPTITCIACNGYK 317
QY 440 WNVLPNGMKTSCFNVGNSKCDGMNGNEVAGDHIOAGAGSDNDYLILNLHIPFKPPTS 499
Db 318 -----PADKTTTCBAVSNCKTPGCKACNEKEVCTDCDSSTYL-----TPTS 362
QY 500 MTGATGSELGRITVFETLCSADCVLYFMVDINRKRSTNVVSWGCTKEKQAYTHIFKNA 559
Db 363 Q-----CIDSCA-----KIGNYGATEGAKK-----ICKEC 388
QY 560 TFFTWAQFQRTNQODNRRFINDMVKIYSITATNAVGVASSCRACALGSEQSSSCVPC 619
Db 389 TA-----ANCKTCDGQ-----QCCQACSDGFYKNGDACSPC 419
QY 620 PPGHYIEK-----ETNOCKECPDPTLYLSIHQVYGEACI--PCGPGSKNNQDHSVYSDCF 673
Db 420 ---HESCKTSAGTASDCTECTPGKALR---YGDGTGKTCGEGCTTGQSGGAC-----467
QY 674 FYHEKENQILHYDFSNLSSVGLMNGPSFTSKGTYKHFNFNISLCGHEGKKMALCTNNIT 733
Db 468 -----KTCGLTIDGASYCECATTEYPONGYV---APKASRATPTCN 507
QY 734 DFTVKEIVAG--SDDYTNLVGAFVQCSTIIPSESKGFRAALSQSIIADTFIGVTVETT 791
Db 508 DSPIQNGVCGTCADNYFKMNGG--CYETV---KYPGKTVCISAPN-----GGTCQKA 554
QY 792 LKNINIKEDMFPVPTSOIPDVHFFYKFSSTATTSCING--RSTAVKMRGN---PTKSGAGV 846
Db 555 ADGKYLDSGLTVCSGCKEC-----TSSTDCTCLDGVKRSASACTKDCSSCETCNGA--608
QY 847 ISVPSKCPAGTCDGC--TFYFLWESAE--ACPLTEHDFHEIEGACKRGFOETLYVNMFPKW 904
Db 609 -----ATTCKACATGYKTAGSAGACATSC--ESDSNGVTGI-----KG 644
QY 905 CIKGISLPEKKLAT--CETVDFWLKVGAGVCAFTAVL-----LVALTCYFW 948
Db 645 CL-NCAPPSSSTGSLVCLYLQNTNKSGLSAGAIAGISVAVIVVAVGLVGLFCLMWF 698

RESULT 8
Q9XTJ7 PRELIMINARY; PRT; 709 AA.
ID Q9XTJ7
AC Q9XTJ7;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Variant-specific surface protein (Fragment).
GN VSP417-6.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiae.
OX NCBI_TaxID=5741;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=BRIS-136;
RX MEDLINE=99026095; PubMed=9806870;
RA Ey P.L., Darby J.M.;
RT "Giardia intestinalis: conservation of the variant-specific surface
RL protein VSP417-1 (TSA417) and identification of a divergent homologue
RL encoded at a duplicated locus in genetic group II isolates."
DR EMBL; U89266; AAD03483.1;
DR InterPro; IPR002174; Furin-like.
```


QY	820	TATTSCINGRSTAVMRCNPTKSGAGVIVSPSKCPAGT-----CDGCTEYFLWESAE-----	871
DR	PROSITE; PS00022; EGF_1; UNKNOWN_11.		
DR	PROSITE; PS01186; EGF_2; 1.		
DR	PROSITE; PS01248; LAMININ_TYPE_EGF; 14.		
FT	NON_TER 2759		
QY	SEQUENCE 2759 AA: 305515 MW: 081F737000C63541 CRC64;		
QY	Query Match	3.5%; Score 199.5; DB 5; Length 2759;	
QY	Best Local Similarity	19.1%; Pred. No. 3.9e-07;	
QY	Matches 225; Conservative 110; Mismatches 363; Indels 483; Gaps 65;		
QY	39 LAGQAAWAGDLPSSSRPLPPC-----QEKDY-----HFEYTECDSSGS 78		
Db	819 LGACEQC---ECPSLDLNPPECISTELAVLGVSASNEQVYVINCPLGYEGNKCEY--- 872		
QY	79 RWRVAIPNSAVDCSG--LPDPVRGK--ECTF-----SCASGEYLEMKNQVSKC- 123		
Db	873 -----CSDGFEDPLGTGKIECTCNGNIDPMGIGNCDS-----ETGKCLKCI 914		
QY	124 GEGTSLGSGTKDEWDELPAFNSIATETMDTVGPGSDSRPGDCNNSWIPRGNIENR 183		
Db	915 GHTTGDSCSEKHHW-----GNAQLT-----CKPGCCHTQGVNQCSENGE 959		
QY	184 DDCVTSIIYAVHLKSGYVFXYVDNNIFFEFTQDCQEMDTTDDKWKLLDNGEW 243		
Db	960 CEKENYIGA-----QCDRCENHGDVNGCPACDCNDT 993		
QY	244 GSHVMLKSGTNILYRTTGILMGSKAVKPVLYKNITIEGVAYTSECFCKPGTSEN--- 300		
Db	994 GS-----IGSDCQ-----VSGQC-NCKQGVFGKQCD 1019		
QY	301 --KPGSRN-----COVPRNTYSEKGAKEIRCKDDSQFSGSECTERPPCT-----TKD 348		
Db	1020 QCRSYFNFTDAGCQFCHNIY---GSIEDGK---DQTGKCEENVEGTMCKEACDG 1073		
QY	349 YFQIHT-----PCDEBEGTKQIMKWIIEPK-TCREDLTDAILRPSGKCKPCPNPGF 400		
Db	1074 YFNITSGDGEDCQDPTGSDVSNLVGTQCVCKPGVT-----GLK--CDSCLPNF 1123		
QY	401 Y---NNGSSSCHPCP-PCTEFSDGTEKCRPCPAGTEPALGFYKWNVLPGNMKTSFCFNVG 456		
Db	1124 YGLTSEGTECECPAPQVCQDPIDGSCVCPNT-----VGEMCENCNT-- 1167		
QY	457 NSKCDGMNGWEVAGDHIOQSGAGSDNDYILNLHLHPGKPPPTSMWATGSELGRITFVFE 516		
Db	1168 -----NAW-----DYHPLN--GCK----- 1179		
QY	517 TLCSADCVLYFNVDINRKNSTNVESWGTKEQAY-----THIFKNATFTFTWAFQ 568		
Db	1180 -LC--DC-----SDIGSDGMCNTFTTGCKCKAAYVGLKDLCTHGFNFPT----- 1223		
QY	569 RTNQGDNRRFINDMVKIYSITATNAVGVASSRACALGSEQSGSCVCPCPGPHYIEKE 628		
Db	1224 -----CEPC--GCNAAGTDPLOCKDQOCL-- 1245		
QY	629 TNQCKECPDPTLYSLTHQVYGRKACIPCPG-----SKNQDHVSVYDCDF----- 673		
Db	1246 CNEIGBCP-----CKKNVHGTR-CDQCGEGTFLSDSNLKGCTECF--CFNRTSCNQSD 1297		
QY	674 -----FYHEKEN-QILHYDFSNLSS-----VGSMLMGPS 701		
Db	1298 LVNQMYAEDRAVTPQEPWEFTYKKNLNLREKPSHNSYPTDATPLYWLPSTMLGDR 1357		
QY	702 FTS-----KGTKYF-HFFNISLCGHEGKKMALCTNNITDFTVK 738		
Db	1358 TASYNGFLRKIWNEDNRRGLHGRPDQOYFRHFPQVIFGNRRIELPHIMEINDDGIY 1417		
QY	739 EI-----VAGSDYT-----NLVGAFCVOSTIIPSESKGFAALSSQSIL 779		
Db	1418 KIRLHESERVRHSPELTUTRKMVMVAQDGTGIIVTYTPARGDAINQEVSLDVAV 1477		
QY	780 ADTFIGVTVEITLKNINIKEDM-----FPVPTSQIPDVHFFYKSS----- 819		
Db	1478 DEKTIAGISTT-KATCVKEKTLCCDQGYNGICSCONPEGYVRKKHREYVNLNOADIALIGW 1536		

Db 873 -----CSDGFEDPLTGKTECTGNGNIDPMGTGNCDS-----ETGKCLKCI 914
Qy 124 GEGTSLGSGIKFDEWDELPAFSGNSIATFMDTVVPSDSRPGDGNSSSWIPRGNYIESNR 183
Db 915 GHTTGDSCSEKHHW-----GNAQLHT-----CKPCCHTQGAVNPQCSERGE 959
Qy 184 DDCVTSIIYAVHLKKSQYFEFYVDNNIFFEFFEFTONOCQOEMDTTDDKWKVLTDNGEW 243
Db 960 CEKENYIGA-----QCDRCKENHGVDENGCFACDCNDT 993
Qy 244 GSHSVMLKSGTNILYWRITGILMSKAVKPVLVKNITIEGVATSECFCKPQTFN--- 300
Db 994 GS-----TGSQCDQ-----VSGQC-NCKQGVFGKQCD 1019
Qy 301 --KPGSFN-----CQVCPRTYSEKAKECIRCKDSDQSGSSECTERPPCT-----TKD 348
Db 1020 QCRPSYFNFTDAGCQFCHCNFY---GSIEDGKC---DQITGKCECRENVGTECKECADG 1073
Qy 349 YFOIHT-----PCDEEGKTQIMYKWIPEK-ICREDLTDAILRPPSGEKKKDCPPCNPGE 400
Db 1074 YFNITSGDGEDCGCDPTGSEDVSCNLVTGQCVCKPGVT-----GLK--CDSCLPNF 1123
Qy 401 Y---NNGSSSCHPCP-PGTFSDGTEKCRPCPAGTEPALGFYKWNVLPGNMKTSFCFNVG 456
Db 1124 YGLTSECTECPAPGQVCPIDGSCVCPNT-----VGEHCENCTT-- 1167
Qy 457 NSKCDGMNGEAVAGDHIQSGAGSDNDYILNLHIFPKPPTSMTGATGSELGRITFVFE 516
Db 1168 -----NAW-----DYHPLN---GCK----- 1179
Qy 517 TLCSADCVLYFMVDINRKNSTNVNVEGKTKEQAY-----THIFKNATFTTMAFQ 568
Db 1180 -LC--DC-----SDIGSDGMCNFTFGQCKCKAAVYGLKCDLCTHGFNFPT----- 1223
Qy 569 RTNQGODNRRFINDMVKIYSITATNAVDGVASSCRACALGSEQSGSCVCPGPHVIEKE 628
Db 1224 -----CEPC---GCNAAGTDPLQCKDGOCL-- 1245
Qy 629 TNCKECPPTVYLSHQVYKKEACIPCGPG-----SKNQDHSVCYSCDF----- 673
Db 1246 CNEIGECP-----CKKNVHGKT-CDQCGEGTFLSDSSNLKGTECF--CFNRTSNCEQSD 1297
Qy 674 -----FYHEREN-OILHYDFSNLSS-----VGLMNGPS 701
Db 1298 LVWQOYAEADRAVQEPPEFTYKKNINLREKPSHFNSYPTDAPLYWPLPSTMLGDR 1357
Qy 702 FTS-----KGTIFY-HFFNISLCGHEGKKMALCTNNITDFTVK 738
Db 1358 TASYNGFLRFKIWNEDNRRLHGIHQDQYPRFPQVIFIGNNRLELHHPMEINDDGIY 1417
Qy 739 EI-----VAGSDDYT-----NLVGAFCVOSTIIPSEKGFRAALSSQSIL 779
Db 1418 KIRLHESEWRVRSPELTLTRKOMMVALQDTQGIYIRGTYYTPARGDAINIQEVSILDVAV 1477
Qy 780 ADTFIGVTETTLKNINIKEDM-----FPVPSQIPDVHFFYKSS----- 819
Db 1478 PESKIVAGLSTT-KAIGVEKLCGCGQGTGLSCQNPVEGYVYRKKHREYLNQADITALIGW 1536
Qy 820 TATTSCINGRSTAVKMRNPTKSGAGVISVPSKCPAGT-----CDGCTFYFLWESAE----- 871
Db 1537 SEPCSC-HGHS-----QTCNPD-----TSVCTDCEHNTFGDFCEHCLPGYIDAREGAN 1585
Qy 872 -----ACPL-----CTEHDHFELEGACKRQFQETLYVWNPFPKWCIG-ISLPEKK 915
Db 1586 ACTKACAPLVENSFSDSCVAVDHGRGYVNSCKPG-----YTGQYCECTCVAGYGDPOHI 1640
Qy 916 LATCETVDFWLKVAGVAGTAVAL-----LVALTCYFWKKKQKLEYKYKSLVMTNS 967
Db 1641 GGTCSPCDCH-PDGLHACAPNLSCQCECKPGVGTCTSCMQE-----RHAFINRVCTSCD 1695
Qy 968 KECELPADASCAIMGEEDNEEYVSNKOSLLGLKLSLATK 1008

Db 1696 QGCYLP-----LMETMDTMEE--HLGRQNFSG-LKPIPWK 1727
RESULT 13
Q8WRD2
ID Q8WRD2 PRELIMINARY: PRT: 3087 AA.
AC Q8WRD2;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Cysteine repeat modular protein 1 pBCRM1.
OS Plasmodium berghei (strain Anka).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5823;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=ANKA;
RC Thompson J.;
RA "Plasmodium berghei Cysteine Repeat Modular Protein 1: pBCRM1.";
RT Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF449196; AAL47156.1; --
DR InterPro: IPR000345; Cytochrome_bind.
DR InterPro: IPR000361; EGF-like.
DR InterPro: IPR001664; IF.
DR InterPro: IPR000001; Kringle.
DR InterPro: IPR000589; Ribosomal_S15.
DR Pfam: PF00051; Kringle; 1.
DR SMART: SM00130; KR; 1.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
DR PROSITE: PS01186; EGF_2; UNKNOWN_3.
DR PROSITE: PS00226; IF; UNKNOWN_1.
DR PROSITE: PS00070; KRINGLE_2; 1.
DR PROSITE: PS00362; RIBOSOMAL_S15; UNKNOWN_1.
SQ SEQUENCE: 3087 AA; 362575 MW; 38BDD25AB65FD43D CRC64;
Query Match 3.5%; Score 197.5; DB 5; Length 3087;
Best Local Similarity 20.0%; Pred. No. 5.7e-07;
Matches 197; Conservative 127; Mismatches 322; Indels 337; Gaps 55;
Qy 102 ECTFSCASGEYLEMKNQVCKGEGTYSILGSGIKFDEWDELPAFSGNSIATFMDTVVGPSPD 161
Db 1317 EFSIACEDGHY--HKKNKCHPCCKEKGYNLNLEIK-----QNN 1351
Qy 162 SRPDGC-----NNSSWIPRGNYIESNRDCTVSLIYAVHLKK-----SGYVFEFYQY 208
Db 1352 ERYKKCTACGQNRITLKKKEFEKN---CLCDLGEY-YIKNPNRPNFICSCVYGEYKD 1407
Qy 209 VDNITFEFFETONOCQOEMDTTDDKWKVLTDNGEWGSHSVMLKSGTNILYWRITGILMS 268
Db 1408 V-----ISNELCK-----GIVCKKNASYL-----ILDK 1430
Qy 269 KAVKPVLVKNITIEGVATSECFCKPQTF--SNKPGSFNCQVCPRTYSEKAKECIR 326
Db 1431 KSVDP-----SOCL-CNGGYLYRDKNNNEICTKINNHY-----C 1465
Qy 327 KDDQSGSSEC-----TERPPCTTKDYFQIHTPCDEGKTQIMYKWIPEKICRE----- 376
Db 1466 PDNDNY--YKCKPIHNKTEEIKRT--DFETINSLCEEGEYEPINNMKIKDTKSRDYHYNN 1521
Qy 377 -----DLTAIRLPPSGEKKKDCPPCNPFGYNNSSS--CHPCPPG--TFSDGTG--- 422
Db 1522 IFVTYKYNLSIV-----NDNNICMECNLFYKKNISSEKCIKCPRSSTTTTYSKSTQN 1576
Qy 423 CRPCPAGTEPALGFYKWNVLPGNMKTS-----PVGNS-KCDGMNGEAVAGDHI- 473
Db 1577 CNSCHKG-----YK-----DQKQVCKCLPNHFCVKGSKQNDKNNISQYAGDAI 1622
Qy 474 -----QSGAGSDNDYILNLH-----IPG 493
Db 1623 CPNYSVTLPQYENNISFKNCLCIKGYEKNFQDFYNNHCKKAPNFYKDTISNDLSIP- 1681
Qy 494 FKPTSTMTGATGSELGRITFVFEFTLCSADCVLYFMVDINRKNSTNVNVEGKTKEKQAYTH 553

Db 1682 -CPTNSITLNTGA-----TSIYNICIDKG---FFYDTVSVCNCPHGY-YCSEKDMTTK 1731
QY 554 I-----IFKNATFTTFAFORTWOGDNRFINDMVKIYSITATNAVGVASSCRACALG 608
Db 1732 LAPPKPKNYTIIYKGSYNISHCVCSGYTANTVIVEHS-SENGIIESHESAKNIKIQ 1790
QY 609 SEQSGSCVPCPGHYTEKTN-QCKECPDPT-----YLSIHQVYGREACI 653
Db 1791 AKNRTSVCVKPCOSSYKSKISNEQCHKCPKNSKTLKDFNNSDIFFCLCTWGYTDKKECK 1850
QY 654 PCG-----PESKNQDHSVCYDCFFYHEKENQILHYDFSNSVSGSLMNGFSFTSGT 707
Db 1851 PCWENKLYCEGEKIYQIETIYDEIINLIQYIRLLSDISTKKSFINIIN--EYINKGL 1908
QY 708 KYHFHFNISCGHGGKMAKLTNNITFTVKEIVAGSDDYT---NLVGAFV-----COS 758
Db 1909 -ISEITHISMMEKSRKSG---NNIQ---SIRKAKKDEIYAVNLYNRLGNIITKRELKT 1962
QY 759 TIIPSESGFRAALSSOSIILADTFIGVTETTL---KNINIKEDMFVPTSQIPDVHFF 815
Db 1963 PKINNOIKIELAKSIENKIIAN---INSKLTLLSEGNKIDTD-----TSIL--ANLF 2011
QY 816 YKSS-----TATTSCIN--GRSTAVMRCNPTKSGAGVISVPS-----KCPAGTCDGCTFYF 865
Db 2012 YKSSNIIYIKHQLINCQRTVIPL-----GVDSSQNFDDCKCKG-----YY 2055
QY 866 L-----WESAACPLCTEHDHELEG-----ACKRGFOETL 896
Db 2056 LEDRVLKNIKICKPCPGTFRKNDGVKKICISCPKSTSIKGSYPNHCFCKNGF---F 2112
QY 897 YVMNEPKWICKIGISLPERKKLATC 919
Db 2113 YSKDTCLELEG-----ATC 2127

RESULT 14
Q99NDO
ID Q99NDO PRELIMINARY: PRT: 5374 AA.
AC Q99NDO;
DT 01-JUN-2001 (TremBLrel. 17, Created)
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE ZAN (Zonadhesin).
GN ZAN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=21138439; PubMed=11239002;
RA Wilson M.D., Riemer C., Martindale D.W., Schnupf P., Boright A.P.,
RA Cheung T.L., Hardy D.M., Schwartz S., Scherer S.W., Tsui L.-C.,
RA Miller W., Koop B.F.;
RT "Comparative analysis of the gene-dense ACHE/TFE2 region on human
RT chromosome 7q22 with the orthologous region on mouse chromosome 5";
RL Nucleic Acids Res. 29:1352-1365(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RA Cheung T.L., Wilson M.D., Koop B.F., Hardy D.M.;
RT "Genomic Basis of Inter- and Intra-species Variation in Zonadhesin
RT Domain Structure";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 3 MAM DOMAINS.
DR EMBL: AF312033; AK28824.1;
DR EMBL: AY046056; AL04416.1;
DR MGD: MGI:106656; Zm.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR003645; FOLN.
DR InterPro: IPR000998; MAM_domain.
DR InterPro: IPR003328; TILA_Cysrich.

DR InterPro: IPR002919; TIL_Cysrich.
DR InterPro: IPR001007; VWF_C.
DR Pfam: PF00629; MAM; 3.
DR Pfam: PF01826; TIL; 25.
DR Pfam: PF02345; TILA; 25.
DR Pfam: PF00094; vwd; 4.
DR SMART: SM00001; EGF_like; 1.
DR SMART: SM00274; FOLN; 21.
DR SMART: SM00137; MAM; 3.
DR SMART: SM00216; VMC; 25.
DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
DR PROSITE: PS0186; EGF_2; 18.
DR PROSITE: PS00060; MAM_2; 3.
KW EGF-like domain; Glycoprotein.
SQ SEQUENCE 5374 AA; 579536 MW; 90D2D8CFE5DE24EB CRC64;

Query Match 3.4%; Score 193.5; DB 11; Length 5374;
Best Local Similarity 19.3%; Pred. No. 2.8e-06;
Matches 233; Conservative 103; Mismatches 387; Indels 487; Gaps 64;

QY 23 RGRSPWPSPAWI-----CCWALAGCQ-----AAWAGDLPSSSR-----PLPPCQE 63
Db 3122 OGALIPGKTIWITSGTQSCNCTGAIQCQNFQCLPKTYCKDLKDGSSNCTNIPL-----Q 3177
QY 64 KDYHFEYTECDSSSRNRVAIPNSAVDCSLG---PDPVRGKECTFSC-ASGEYLEMKNQ- 118
Db 3178 CPAHSRYTNCPLS-----CPPSLDPEGLCEGTSKVPSTCREGICQPGYLMHNKC 3230
QY 119 -----VCS-----KCGEYVSLSGIKFDE 138
Db 3231 VLRIFCGCKNTQGAIFISADKTIWISRGCTQSCSAGAIHCNFKPCSGTCKNGDNGSSN 3290
QY 139 WDELPAESNATFMTDVTWVGPDSRDPDGCNNSSWIPRGNYTESNRDCTVSLIYAVHLKK 198
Db 3291 CTEITLQCPNTSQFTDCL---PSCVPSCNRCEVTSVPVSSCREG-----LCN 3337
QY 199 SGYVFFEYQYVNNIFFEFFFIONQCEMDTTTDKWKVLTONGENGSHVLMKSGTNIY 258
Db 3338 HGFVFSE-----DKCVPTCGCKDARGAIIIPAG----- 3366
QY 259 WRTTGILMGSKAVKPVLVKNITIEGVATSECFCKPCKPTFSNKGPFNCQ--VCPRNITYS 316
Db 3367 -----KTWTSKGCTQSCACV-----EGNIQCQNFQCPPEYV- 3397
QY 317 EKGAKECIRCKDDSQFSSESCTE---RPCTTKDYFQIHTPC-----DEEGKTQI 364
Db 3398 -----CKDNSE--GSSTCTKITLQCPAHTQ-----YTSLPCLPCLDPEG----- 3437
QY 365 MYKWTEPKI---CRE-----DLTDAI--RLPPSGEKKDCCPPCNP 399
Db 3438 LCKDTSIPKVPSTCKEGVCQSGYVLNSDKCVLRABECDCKDAQCALIPAGKTWTSPGCTQS 3497
QY 400 -FYNNSSSCH--PCPPGTF---SDGTHKECR-----PCPA----- 428
Db 3498 CACMGAVQCCSQSCPPGTYCKDNEDGNSNCARITLQCPAHSLEFNTCLPCLPCLDPPG 3557
QY 429 ---GTEPALGFYKWNVLPGNNKTSC---FNVGNSKCDGMN----- 464
Db 3558 LCKGASPK-----VPSTCKECCICQSGYVLNSNKKLLNRNCGCKDAHCALLIPDK 3607
QY 465 -----GWEV-----AGDHIOGAGGSSNDYLI-----LNLHI----- 491
Db 3608 TWVSRGCTQSCVCTGGSIQCLSFQCPPGAYCKDNEDGSSNCARIPQCPANSHYTCFPP 3667
QY 492 --PGFKPPTSMTCATSELSGRITTFEETLCSADCVLY--FMVDINKKSTNVNWSWGCTKE 547
Db 3668 CPPSCSDPEGHCEASGRV-----PSTCRECLNCPGEVLD-RDKCYPRVEC--GCKD 3717
QY 548 KQAVTHIIFKNATFTTFAFQ-----RTNQ-----QODNRRFINDMVKI--YS 588
Db 3718 AQG--ALIPSGKTIWISPTQSCACMGVVCQSSQCPPTGTYCKDNEDGNSNCARITLQ 3775

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 00:12:42 ; Search time 4334.16 Seconds
(without alignments)
13287.722 Million cell updates/sec

Title: US-10-073-333A-3
Perfect score: 3536
Sequence: 1 atgctgttccgcgcgcggg.....ggtaaaaaaaaaaaaaa 3556

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:

- 1: em_estba:**
- 2: em_esthum:**
- 3: em_estin:**
- 4: em_estmu:**
- 5: em_estov:**
- 6: em_estpl:**
- 7: em_estro:**
- 8: em_htc:**
- 9: gb_est1:**
- 10: gb_est2:**
- 11: gb_htc:**
- 12: gb_est3:**
- 13: gb_est4:**
- 14: gb_est5:**
- 15: em_estfun:**
- 16: em_estom:**
- 17: gb_gss:**
- 18: em_gss_hum:**
- 19: em_gss_inv:**
- 20: em_gss_pln:**
- 21: em_gss_vrt:**
- 22: em_gss_fun:**
- 23: em_gss_mam:**
- 24: em_gss_mus:**
- 25: em_gss_other:**
- 26: em_gss_pro:**
- 27: em_gss_rod:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	704	19.8	705	BQ182610	BQ182610 UI-H-EUO-
2	660.4	18.6	697	BM723980	BM723980 UI-E-EOL-
3	654.4	18.4	656	BM783507	BM783507 K-EST0061
4	651.2	18.3	713	BQ772508	BQ772508 UI-H-EUO-
5	624.8	17.6	930	BF035563	BF035563 601457452
6	610.4	17.2	646	AW954806	AW954806 EST366876

7	584.4	16.4	896	12	BG116864	BG116864 602318965
8	553.2	15.6	762	13	BI227132	BI227132 602952315
9	527	14.8	527	14	BM752466	BM752466 K-EST0029
10	513.4	14.4	667	10	BB384282	BB384282 BB384282
11	512.8	14.4	634	14	BQ551364	BQ551364 H4008H01-
12	507.4	14.3	515	14	BM771726	BM771726 K-EST0055
13	504	14.2	576	14	W60844	W60844 zc26f10.sl
14	498.4	14.0	625	12	BG724280	BG724280 602697903
15	494.2	13.9	801	13	BI731892	BI731892 603353640
16	493	13.9	541	10	BE237235	BE237235 146538 MA
17	483	13.6	483	14	BM759885	BM759885 K-EST0040
18	482.2	13.6	557	14	W60843	W60843 zc26f10.r1
19	481.4	13.5	484	9	AA010992	AA010992 ze21f02.r
20	.478	13.4	478	10	AW665873	AW665873 h195a04.x
21	477	13.4	477	9	AI885785	AI885785 w161e05.x
22	472.6	13.3	585	14	BM974911	BM974911 UI-CF-EC1
23	470	13.2	1773	11	AK017241	AK017241 Mus muscu
24	468.8	13.2	718	12	BG086402	BG086402 H3125F06-
25	449.4	12.6	508	13	BM538241	BM538241 ha32b12.g
26	446	12.5	458	12	BF740039	BF740039 7042f10.x
27	435.6	12.2	1109	14	BQ896007	BQ896007 AGENCOURT
28	424.8	11.9	600	10	BB617824	BB617824 BB617824
29	421.8	11.9	656	10	BB626387	BB626387 BB626387
30	419.2	11.8	653	10	BB650824	BB650824 BB650824
31	414.4	11.7	507	9	AI797353	AI797353 we87b06.x
32	412.4	11.6	565	13	BI337511	BI337511 360177 MA
33	404	11.4	420	13	BI480715	BI480715 H2RPE-023
34	397	11.2	445	9	AA010869	AA010869 ze21f02.s
35	394.4	11.1	641	10	BB629056	BB629056 BB629056
36	391.8	11.0	667	13	BM491171	BM491171 pqp20.pk0
37	388.6	10.9	640	10	BB622295	BB622295 BB622295
38	381.8	10.7	859	14	BQ730700	BQ730700 AGENCOURT
39	374.6	10.5	449	14	HI8553	HI8553 ym43d01.r1
40	374	10.5	374	9	AA644099	AA644099 ab62c03.s
41	370.8	10.4	446	9	AI317059	AI317059 uk71h12.y
42	369.4	10.4	395	10	AW444960	AW444960 UI-H-BI3-
43	344.4	9.7	553	10	BB644517	BB644517 BB644517
44	343.8	9.7	631	10	AV335504	AV335504 AV335504
45	338	9.5	603	13	BJ081522	BJ081522 BJ081522

ALIGNMENTS

RESULT 1 BQ182610/c BQ182610 705 bp mrna linear EST 30-APR-2002
LOCUS UI-H-EUO-azz-a-23-0-UI.s1 NCI-CGAP_Carl Homo sapiens cDNA clone
DEFINITION IMAGE: 5854750 3', mRNA sequence.
ACCESSION BQ182610
VERSION BQ182610.1 GI:20358160
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 705)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
The following repetitive elements were found in this cDNA
sequence: 1-40, >AT-rich#low_complexity (matched complement)
Seq primer: M13 FORWARD
POLYA=Yes.

REFERENCE BQ182610
AUTHORS
TITLE
JOURNAL
COMMENT

FEATURES		Location/Qualifiers		source	
1. 705		/organism="Homo sapiens"		/db_xref="taxon:9606"	
		/clone="IMAGE: 5854750"		/clone_lib="NCI_CGAP_Carl"	
		/tissue_type="Osteoarthritic Cartilage"		/dev_stage="Adult"	
		/lab_host="DH10B (Life Technologies)"		/note="Organ: Knee; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site.1: EcoR I; Site.2: Not I; NCI_CGAP_Carl is a cDNA library containing the following tissue(s): Osteoarthritic Cartilage The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is TGATCAGCT."	
BASE COUNT		194 a	136 c	118 g	257 t
ORIGIN					
Query Match		19.8%; Score 704; DB 14; Length 705;			
Best Local Similarity		100.0%; Pred. No. 1.3e-131;			
Matches 704; Conservative		0; Mismatches 0; Indels 0; Gaps 0;			
QY	2853	TCAAAACCTGGAATACAAATATCCAACTAGTATGAGCTAACTCAAAAGAGCTGTGA	2912		
DB	705	TCAAAACCTGGAATACAAATATCCAACTAGTATGAGCTAACTCAAAAGAGCTGTGA	646		
QY	2913	ACTCCGGCTGCAGACAGTTGCTATCATGAAGGAGAGATATGAAGGAAGTTGT	2972		
DB	645	ACTCCGGCTGCAGACAGTTGCTATCATGAAGGAGAGATATGAAGGAAGTTGT	586		
QY	2973	ATATTCCAATAACAGCTCACTACTAGGAACCTCAAACTCTTGGCAACCAAGAAAAAGA	3032		
DB	585	ATATTCCAATAACAGCTCACTACTAGGAACCTCAAACTCTTGGCAACCAAGAAAAAGA	526		
QY	3033	AGACCATTTTGAATCTGTTCAACTGAAACCTCAAGATCCCCCAATATATGAAGAGACAG	3092		
DB	525	AGACCATTTTGAATCTGTTCAACTGAAACCTCAAGATCCCCCAATATATGAAGAGACAG	466		
QY	3093	TGCTGTAGCCTTGAGACTAATGAACAAAGAACCTGCTCTAGTTTACAGGACCATATTT	3152		
DB	465	TGCTGTAGCCTTGAGACTAATGAACAAAGAACCTGCTCTAGTTTACAGGACCATATTT	406		
QY	3153	TAGGGTCTGTCCTCATACCTGTCACATTCGTCATCTCACAGAGGGGCCATGCCGCTGA	3212		
DB	405	TAGGGTCTGTCCTCATACCTGTCACATTCGTCATCTCACAGAGGGGCCATGCCGCTGA	346		
QY	3213	AAAGGAAGGAGATTGAACATTTGATTCGCTTATCAGATGTCACAGTACCTTGCCAAAT	3272		
DB	345	AAAGGAAGGAGATTGAACATTTGATTCGCTTATCAGATGTCACAGTACCTTGCCAAAT	286		
QY	3273	AAAGGAAGCAATGATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAGGAAGAGATTT	3332		
DB	285	AAAGGAAGCAATGATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAGGAAGAGATTT	226		
QY	3333	ATCTGTATATACATCAACTGAAACCAAGTTTAAAGCCCAACCACTGCCTGATGCAT	3392		
DB	225	ATCTGTATATACATCAACTGAAACCAAGTTTAAAGCCCAACCACTGCCTGATGCAT	166		
QY	3393	GCCATATAATTAATGGTGAACCTTTTATTCCTTTATGATGTCTACATAACAGTGTGATTTG	3452		
DB	165	GCCATATAATTAATGGTGAACCTTTTATTCCTTTATGATGTCTACATAACAGTGTGATTTG	106		
QY 3453		GAAGGCACATGTGAGCATATGACATATGATCCATTTATGTTTCTTTGTTTATATTTT 3512			
DB 105		GAAGGCACATGTGAGCATATGACATATGATCCATTTATGTTTCTTTGTTTATATTTT 46			
QY 3513		TGGGGAATTAATAATTTTTTAAAGCTAAAAAATAAAAAAATAAAAAA 3556			
DB 45		TGGGGAATTAATAATTTTTTAAAGCTAAAAAATAAAAAAATAAAAAA 2			
RESULT 2					
LOCUS		BM723980			
DEFINITION		UI-E-E01-aiy-a-05-0-UI.r1 UI-E-E01 Homo sapiens cDNA clone			
ACCESSION		BM723980			
VERSION		BM723980.1 GI:19045304			
KEYWORDS		EST.			
SOURCE		human.			
ORGANISM		Homo sapiens			
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS		1 (bases 1 to 697)			
TITLE		Bonaldo,M.F., Lennon,G. and Soares,M.B.			
JOURNAL		Normalization and subtraction: two approaches to facilitate gene discovery			
MEDLINE		Genome Res. 6 (9), 791-806 (1996)			
COMMENT		97044477 Contact: Soares, MB Program for Rat Gene Discovery and Mapping University of Iowa 451 Eckstein Medical Research Building Iowa City, IA 52242, USA Tel: 319 335 8250 Fax: 319 335 9565 Email: msoares@blue.weeg.uiowa.edu Tissue Procurement: Dr. Gregg Hageman cDNA Library preparation: Dr. M. Bento Soares, University of Iowa cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa cDNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com). Seq primer: M13 Reverse.			
FEATURES		Location/Qualifiers			
source		1. 697 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="UI-E-E01-aiy-a-05-0-UI" /clone_lib="UI-E-E01" /tissue_type="fetal eye" /dev_stage="fetal" /lab_host="DH10B (Life Technologies) (T1 phage resistant)" /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site.1: EcoR I; Site.2: Not I; UI-E-E01 is a normalized cDNA library containing the following tissue(s): fetal eye. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CCGGTATACC. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."			
BASE COUNT		222 a	150 c	138 g	186 t
ORIGIN					
Query Match		18.6%; Score 660.4; DB 14; Length 697;			
Best Local Similarity		99.4%; Pred. No. 8e-123;			
Matches 694; Conservative		0; Mismatches 1; Indels 3; Gaps 3;			

```
QY 1835 CGGGTTTCATCGTGTGTCCTCCCTCCAGGCCACTACAT-TGAGAAAGAAACCAACACAG 1893
Db 1 CGGGTTTCATCGTGTGTCCTCCCTCCAGGCCACTACATNTGAGAAAGAAACCAACACAG 60
QY 1894 TGAAGGAATGCCACCTGACACCTACCTGTCATACATCAGGTCTATGCGAAAGAGGCT 1953
Db 61 TGAAGGAATGCCACCTGACACCTACCTGTCATACATCAGGTCTATGCGAAAGAGGCT 120
QY 1954 TGTATTCCATCGGGGCTGGGAGTAAAAACAATCAGGACCATTCGGTTTGGCTATAGTGAC 2013
Db 121 TGTATTCCATCGGGGCTGGGAGTAAAAACAATCAGGACCATTCGGTTTGGCTATAGTGAC 180
QY 2014 TGCCTTTTCTACCATCAAAAGAAATCAGATTTTGCACATATGACATTTAGCAACCTCAGC 2073
Db 181 TGCCTTTTCTACCATCAAAAGAAATCAGATTTTGCACATATGACATTTAGCAACCTCAGC 240
QY 2074 AGTGTGGGCTCATTATGAATGCCGCCAGCTTCACCTCCAAAGGAACAAATACATTCAT 2133
Db 241 AGTGTGGGCTCATTATGAATGCCGCCAGCTTCACCTCCAAAGGAACAAATACATTCAT 300
QY 2134 TTCTTCAATATCAGTTTATGTGGGATGAGGGAAGAGATGGCTCTCTGTACCAACAAT 2193
Db 301 TTCTTCAATATCAGTTTATGTGGGATGAGGGAAGAGATGGCTCTCTGTACCAACAAT 360
QY 2194 ATAACAGACTTTACAGTAAAGAAATAGTGGCAGGTGAGATGATTACACAAA-TTTGGT 2252
Db 361 ATAACAGACTTTACAGTAAAGAAATAGTGGCAGGTGAGATGATTACACAAA-TTTGGT 420
QY 2253 AGGGGCAATTTGATGCGAGTCAACAATATTCTTCTGAAAGTAAGGTTTCCGAGCAGC 2312
Db 421 AGGGGCAATTTGATGCGAGTCAACAATATTCTTCTGAAAGTAAGGTTTCCGAGCAGC 480
QY 2313 CTTATCATCAATCCATCATTTCTGGCAGATACATTTATGAGGTGACAGTTTGAACACAC 2372
Db 481 CTTATCATCAATCCATCATTTCTGGCAGATACATTTATGAGGTGACAGTTTGAACACAC 540
QY 2373 ATTGAAATATATATAAAGAGATATGTTCCAGTTTCCAAAGCCAAATACACAGA 2432
Db 541 ATTGAAATATATATAAAGAGATATGTTCCAGTTTCCAAAGCCAAATACACAGA 600
QY 2433 TGTGATTTCTTTTATAAGTCTTTACAGCAACAACATCTTCTATTATGCGCGATCAAC 2492
Db 601 TGTGATTTCTTTTATAAGTCTTTACAGCAA-AACATCTTGTATTATGCGCGATCAAC 659
QY 2493 TGTCTGAAATGAGGTGTAATCTTACTAAATCTGGAG 2530
Db 660 TGTCTGAAATGAGGTGTAATCTTACTAAATCTGGAG 697

RESULT 3
BN783507
LOCUS K-EST0061445 S7SNU719 Homo sapiens cDNA clone S7SNU719-36-H08 5',
DEFINITION mRNA sequence.
ACCESSION BN783507
VERSION BN783507.1 GI:19131739
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 656)
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
21C Frontier Korean EST Project 2001
Unpublished (2002)
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel.: +82-42-860-4470
Fax: +82-42-860-4409
```

FEATURES
source

Email: yongsung@mail.kribb.re.kr
Plate: 36 row: H column: 08
High quality sequence stop: 656.
Location/Qualifiers
1..656

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S7SNU719-36-H08"
/clone_lib="S7SNU719"
/sex="M"
/tissue_type="Stomach"
/cell_type="Epithelial"
/cell_line="SNU-719"
/lab_host="Top10F"
/note="Organ: Stomach; Vector: pcNS; Site_1: EcoRI;
Site_2: NdeI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tobacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."

BASE COUNT 191 a 147 c 145 g 173 t
ORIGIN

Query Match 18.4%; Score 654.4; DB 14; Length 656;
Best Local Similarity 99.8%; Pred. No. 1.3e-121;
Matches 655; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1710 TAATCAGGCTCAAGATAATAGACGGTTTCATCAATGACATGGTGAAGATTTATCTATCAC 1769
Db 1 TAATCAGGCTCAAGATAATAGACGGTTTCATCAATGACATGGTGAAGATTTATCTATCAC 60

QY 1770 AGCCACTAATCAGTTGATGGGTGGCGTCCCTCATCGCGTGGCTGTGCCCTCGGTTCTGA 1829
Db 61 AGCCACTAATCAGTTGATGGGTGGCGTCCCTCATCGCGTGGCTGTGCCCTCGGTTCTGA 120

QY 1830 ACAGTCGGGTTTCATCGTGTGCCCTCCAGGCCACTACATTTGAGAAAGAACCAA 1889
Db 121 ACAGTCGGGTTTCATCGTGTGCCCTCCAGGCCACTACATTTGAGAAAGAACCAA 180

QY 1890 CCAGTGAAGGAATGTCCACCTGACACCTTACCTGTCCATACATCAGGTCTATGGCAAGA 1949
Db 181 CCAGTGAAGGAATGTCCACCTGACACCTTACCTGTCCATACATCAGGTCTATGGCAAGA 240

QY 1950 GCGTTGTATTCCATCGGGGCTGGGAGTAAAAACAATCAGGACCATTCGGTTTGTATAG 2009
Db 241 GCGTTGTATTCCATCGGGGCTGGGAGTAAAAACAATCAGGACCATTCGGTTTGTATAG 300

QY 2010 TGACTGCTTTTCTTACCATGAAAGAAATCAGATTTTGCACATATGACTTTAGCAACT 2069
Db 301 TGACTGCTTTTCTTACCATGAAAGAAATCAGATTTTGCACATATGACTTTAGCAACT 360

QY 2070 CAGCAGTGTGGGCTCATTAAATGAATGGCCCTTACCTCCAAAGGAACAAATACCTT 2129
Db 361 CAGCAGTGTGGGCTCATTAAATGAATGGCCCTTACCTCCAAAGGAACAAATACCTT 420

QY 2130 CCATTTCTTCAATATCAGTTTATGTGGGATGAGGGGAAGAGATGGCTCTCTGTACCAA 2189
Db 421 CCATTTCTTCAATATCAGTTTATGTGGGATGAGGGGAAGAGATGGCTCTCTGTACCAA 480

QY 2190 CAATATACAGACTTTACAGTAAAGAAATAGTGGCAGGTGAGATGATTACACAAATTT 2249
Db 481 CAATATACAGACTTTACAGTAAAGAAATAGTGGCAGGTGAGATGATTACACAAATTT 540

```
QY 2250 GGTAGGGCCATCTGATCCAGTCAACAATATTCTCTGAAAGTAAGGGTTCCGAGC 2309
Db 541 GGTAGGGCCATCTGATCCAGTCAACAATATTCTCTGAAAGTAAGGGTTCCGAGC 600
QY 2310 AGCTTATCATCATCATCATCTGCGAGATACATATTCATAGGAGTCACAGTTG 2365'
Db 601 AGCTTATCATCATCATCATCTGCGAGATACATATTCATAGGAGTCACAGTTG 656

RESULT 4
BQ772508/c 713 bp mRNA linear EST 26-JUL-2002
LOCUS UI-H-FEO-bbq-b-01-0-UI.s1 NCI_CGAP_FEO Homo sapiens cDNA clone
DEFINITION UI-H-FEO-bbq-b-01-0-UI 3', mRNA sequence.
ACCESSION BQ772508
VERSION BQ772508.1 GI:21980984
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 713)
REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: James Martin
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu
The following repetitive elements were found in this cDNA
sequence: 1-40, >AT-rich#Low_complexity (matched complement)
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES
Source
Location/Qualifiers
1..713
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-H-FEO-bbq-b-01-0-UI"
/clone_lib="NCI_CGAP_FEO"
/tissue_type="Chondrosarcoma Cell line"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT73-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoR I; Site_2: Not I; NCI_CGAP_FEO is
a cDNA library containing the following tissue(s): a pool
of 3 chondrosarcoma cell lines (grade 2) The library was
constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT73-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
CCTACGGAC. The cell lines was provided by Dr James Martin
of University of Iowa.
TAG_LIB=UI-H-FEO
TAG_TISSUE=Human grade 2 chondrosarcoma cell line pool
TAG_SEQ=CGCTACGGAC"

BASE COUNT 193 a 136 c 120 g 258 t 6 others
ORIGIN
Query Match 18.3%; Score 651.2; DB 14; Length 713;
Best Local Similarity 98.2%; Pred. No. 5.7e-121;
Matches 699; Conservative 0; Mismatches 7; Indels 6; Gaps 4;
QY 2851 ATCAAAACTGGAAATACAAATATTCACGTTAGTACGACGACTACCTCA-
AAAGAGTG 2909
```

```
Db 713 AATCAANAACTGGAATACAAATATNCCAAGTTAGTATGACGACTAACTCANAAAGAGTG 634
QY 2910 TGAACATCCCGGCTGCAGACAGTTTGTCTATCATGGAAGGAGAGATTAATGAAGAGAGT 2969
Db 653 TGAACATCCCGGCTGCAGACAGTTTGTCTATCATGGAAGGAGAGATTAATGAAGAGAGT 594
QY 2970 TGTATATTCCAATTAACAGCTCACTACTAGGAAACCTCAAAATCTTTGGCAACCAAGGAAAA 3029
Db 593 TGTATATTCCAATTAACAGCTCACTACTAGGAAACCTCAAAATCTTTGGCAACCAAGGAAAA 534
QY 3030 AGAAGACCATTTTGAATCTGTTCAACTGAAACCTCAAGATCCCAAAATATATGAGAGAGA 3089
Db 533 AGAAGACCATTTTGAATCTGTTCAACTGAAACCTCAAGATCCCAAAATATATGAGAGAGA 474
QY 3090 CAGTGTCTGTAGCCCTTGAGACTTAATGAACAAAGAAACCTGCTCTAGTTTACAGAGACCATA 3149
Db 473 CAGTGTCTGTAGCCCTTGAGACTTAATGAACAAAGAAACCTGCTCTAGTTTACAGAGACCATA 414
QY 3150 TTTTAGGGTCTGTCTCTCATACCTGTCACATTTGGTGATCTCACAGAGGAGGCCCATGCCG 3209
Db 413 TTTTAGGGTCTGTCTCTCATACCTGTCACATTTGGTGATCTCACAGAGGAGGCCCATGCCG 354
QY 3210 TGAAGAGGAGGAGATTGAAACATTTGATTGCCCTTATCAGATGTTCAAGTACCTTGCCA 3269
Db 353 TGAAGAGGAGGAGATTGAAACATTTGATTGCCCTTATCAGATGTTCAAGTACCTTGCCA 294
QY 3270 AATAAGGAAAGCAAAATGATTGGTCTCAACTGAAGATGAAGCTCACTCAGGAGAGA 3329
Db 293 AATAAGGAAAGCAAAATGATTGGTCTCAACTGAAGATGAAGCTCACTCAGGAGAGA 234
QY 3330 TTTATCTGTATATACATACTGAAACCAAGTTTAAAGCCCAACCAATGCCTGCTGATG 3389
Db 233 TTTATCTGTATATACATACTGAAACCAAGTTTAAAGCCCAACCAATGCCTGCTGATG 174
QY 3390 CATGCCATATAATTAATGGGTAACCTTTTATCTTTATGATGTCTACATAA-CAAGTGTGA 3448
Db 173 CATGCCATATAATTAATGGGTAACCTTTTATCTTTATGATGTCTACATAA-CAAGTGTGA 114
QY 3449 TTTGGAAGGCACA-TGTGAGCATATGCATTAT--GATCCAATTTATGTTTCTTTCTTGT 3504
Db 113 TTTGGAAGGCACANTGTGAGCATATGCATTNNATGATCCCAATTTATGTTTCTTTCTTGT 54
QY 3505 TTATATTTTGGGAAAAATTAATAATTTTAAAGTAAAAAATAAAAAAAAAAAAA 3556
Db 53 TTATATTTTGGGAAAAATTAATAATTTTAAAGTAAAAAATAAAAAAAAAAAAA 2

RESULT 5
BQ735563
LOCUS 601457452F1 NIH_MGC_66 Homo sapiens cDNA clone IMAGE:3861156 5',
DEFINITION mRNA sequence.
ACCESSION BQ735563
VERSION BQ735563.1 GI:10743303
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 930)
REFERENCE NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTp
cDNA Library preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
```

```
Plate: LLM9597 row: 9 column: 13
High quality sequence stop: 647.
Location/Qualifiers
1. .930
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3861156"
/clone_lib="NIH_MGC_66"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="organ: ovary; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."
BASE COUNT 275 a 222 c 211 g 222 t
ORIGIN

Query Match 17.6%; Score 624.8; DB 12; Length 930;
Best Local Similarity 91.4%; Pred. No. 1.1e-115;
Matches 684; Conservative 0; Mismatches 62; Indels 2; Gaps 2;

QY 1661 TCATCTTCAAGATGCAACTTTTACATTTGATGGCATTCCAGAGAACTAATCAGGGTC 1720
DB 1 TCATCTTCAAGATGCAACTTTTACATTTGATGGCATTCCAGAGAACTAATCAGGGTC 60

QY 1721 AAGATAATAGACGGTTTCATCAATGATGATGGTGAAGATTTTATCTATCAGGCCACTAATG 1780
DB 61 AAGATAATAGACGGTTTCATCAATGATGATGGTGAAGATTTTATCTATCAGGCCACTAATG 120

QY 1781 CAGTTGATGGGTGGCGTCTCTATGCGGTGCTGCGCTCGGTCGAGAGTGGGTT 1840
DB 121 CAGTTGATGGGTGGCGTCTCTATGCGGTGCTGCGCTCGGTCGAGAGTGGGTT 180

QY 1841 CATCGTGTCTCCCTGCGCTCCAGGCCACTACATTTGAGAGAAACCAACCAAGTCAAGG 1900
DB 181 CATCGTGTCTCCCTGCGCTCCAGGCCACTACATTTGAGAGAAACCAACCAAGTCAAGG 240

QY 1901 AATGTCACCTGACACCTACCTGTGTCATACATCAGGTCTGTATGCAAGAGGCTTGATTC 1960
DB 241 AATGTCACCTGACACCTACCTGTGTCATACATCAGGTCTGTATGCAAGAGGCTTGATTC 300

QY 1961 CATGGGGCTGGAGTAAACAACTACAGGACCATTCGGTTTGGCTATAGTACTGCTTTT 2020
DB 301 CATGGGGCTGGAGTAAACAACTACAGGACCATTCGGTTTGGCTATAGTACTGCTTTT 360

QY 2021 TCTACCATGAAAGAAATACAGATTTTGCACATGACCTTTAGCAACCTCAGCAGTGTGG 2080
DB 361 TCTACCATGAAAGAAATACAGATTTTGCACATGACCTTTAGCAACCTCAGCAGTGTGG 420

QY 2081 GCTCATTAATGATGGCCCGCAGCTTCACCTCCAAAGGAACAAATATCTTCCATTTTCA 2140
DB 421 GCTCATTAATGATGGCCCGCAGCTTCACCTCCAAAGGAACAAATATCTTCCATTTTCA 480

QY 2141 ATATCAGTTTATGTGGCATGAGGGGAGAGAGATGGCTCTCTGTACCAACATATAACAG 2200
DB 481 ATATCAGTTTATGTGGCATGAGGGGAGAGAGATGGCTCTCTGTACCAACATATAACAG 540

QY 2201 ACTTTACAGTAAAGAAATAGTGGCAGGTCAGATGATTACAAATTTGGTAGGGGCAT 2260
DB 541 ACTTTACAGTAAAGAAATAGTGGCAGGTCAGATGATTACAAATTTGGTAGGGGCAT 599

QY 2261 TTGTATGCCAGTCACAATATTCTCTGAAAGTAAGGTTTCCGAGCAGGCTTATCAT 2320
DB 600 TTGTATGCCAGTCACAATATTCTCTGAAAGTAAGGTTTCCGAGCAGGCTTATCAT 659

QY 2321 CACAATCCA-TCAATCTGCGACATCATTTAGAGGTGACAGTTGAACACCATTTGAA 2379
DB 660 CACAATCCA-TCAATCTGCGGCGATCATTTCTAGGATGACAGGTTGAACACCATTTG 719

QY 2380 AATATTAAATATAAAGAGATATGTTCC 2407
DB 720 AACTTTTATTTACAGAGATATGTTCC 747
```

```
RESULT 6
AW954806
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
BASE COUNT
ORIGIN
Query Match
Best Local Similarity
Matches
Conservative
0;
Mismatches
6;
Indels
2;
Gaps
2;
QY 1440
DB 1
QY 1500
DB 61
QY 1560
DB 121
QY 1620
DB 181
QY 1680
DB 241
QY 1740
DB 301
QY 1800
DB 361
QY 1860
DB 421
QY 1920
```

```
AW954806
EST366876
MAGE
Homo sapiens
cdna, mRNA
linear
EST 01-JUN-2000
MAGE
Homo sapiens
cdna, mRNA
sequence.
AW954806
GI:8144489
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspard, R., Gay, C., Holt
I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeaman, T.J. and
Quackenbush, J.
Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
Unpublished (2000)
Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 69
Seq primer: Reverse.
location/Qualifiers
1. .646
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequencing, MAGC"
/note="Vector: pBluescriptSkm"
BASE COUNT 189 a 143 c 142 g 171 t 1 others
ORIGIN
Query Match 17.2%; Score 610.4; DB 10; Length 646;
Best Local Similarity 98.8%; Pred. No. 9.7e-113;
Matches 636; Conservative 0; Mismatches 6; Indels 2; Gaps 2;
QY 1440
DB 1
QY 1500
DB 61
QY 1560
DB 121
QY 1620
DB 181
QY 1680
DB 241
QY 1740
DB 301
QY 1800
DB 361
QY 1860
DB 421
QY 1920
```

|||||

Db 481 CCTGTCCATACATCAGTCTATGCAAGAGGCTTGTTATTCATCGCGGCTGGAGTAA 540

QY 1980 AAACAATCAGACACCATCGGTTTCTATAGTACTGCTGCTTTTCTTACCATG-AAAAGAAA 2038

|||||

Db 541 AAACAATCAGACACCATCGGTTTCTATAGTACTGCTGCTTTTCTTACCATGAAAAGAAA 600

QY 2039 ATCAGATTTGCATATGAC-TTTAGAACCTCAGCAGTGTGGG 2081

|||||

Db 601 ATCAGAGTTTGCATATGACNTTTGGCAACCTTAACAGGTTGGG 644

RESULT 7

BGI16864 896 bp mRNA linear EST 30-JAN-2001

DEFINITION 602318965F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4423217 5', mRNA sequence.

ACCESSION BGI16864

VERSION BGI16864.1 GI:12610303

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 896)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L16M10165 row: b column: 18
High quality sequence stop: 740.
Location/Qualifiers

FEATURES

1. 896

Source

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4423217"

/clone_lib="NIH_MGC_88"

/tissue_type="duodenal adenocarcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: small intestine; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dr primed. Average insert size 1.767 Kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library."

BASE COUNT 299 a 145 c 214 g 238 t

ORIGIN

Query Match 16.4%; Score 584.4; DB 12; Length 896;

Best Local Similarity 97.4%; Pred. No. 1.5e-107;

Matches 637; Conservative 0; Mismatches 11; Indels 6; Gaps 4;

QY 2889 GACGACTACTCAAAAGAGTGTGAACCTCCGCTGCAGACAGTTGCTATCATGGAAGG 2948

|||||

Db 1 GACGACTACTCAAAAGAGTGTGAACCTCCGCTGCAGACAGTTGCTATCATGGAAGG 60

QY 2949 AGAAGATAATGAAGAGGAAGTTGTATATTCCAATAACAGTCACTACTAGGAAAACTCAA 3008

|||||

Db 61 AGAGATATATGAAGAGAGTTGTATATTCCTAATAACAGTCACTACTAGGAAAACTCAA 120

QY 3009 ATCTTTGGCAACCAAGAAAAGAGACCATTTTGAATCTGTTCACTGAAACCTCAAG 3068

|||||

Db 121 ATC-TTGGCAACCAAGAAAAGAGACCA-TTTGAATCTGTTCACTGAAACCTCAAG 178

QY 3069 ATCCCAATATATGAAGACAGTGTGCTAGCCTTGAGACTAATCAACAAAGAACCTG 3128

|||||

179 ATCCCAATATATGAAGAGACAGTGTAGCCTTGAGACTAATGAACAAAGAACCTG 238

QY 3129 CTCTAGTTTACAGGACCATATTTAGGGTCTGCTCATACCTGTCACATTTGGTGTATCT 3188

|||||

Db 239 CTCTAGTTTACAGGACCATATTTAGGGTCTGCTCATACCTGTCACATTTGGTGTATCT 298

QY 3189 CACAGAGGAGGCCATGCCGCTGAAAAGGAGGAGATTTGAACATTTGATTTGCTTATC 3248

|||||

Db 299 CACAGAGGAGGCCATGCCGCTGAAAAGGAGGAGATTTGAACATTTGATTTGCTTATC 358

QY 3249 ACATGGTCAAGTACCTTTGCCAAATTAAGGAAAGCAATGATTTGGGTCTCAACTGAAGAT 3308

|||||

Db 359 ACATGGTCAAGTACCTTTGCCAAATTAAGGAAAGCAATGATTTGGGTCTCAACTGAAGAT 418

QY 3309 GAAGCTCAACTCAGAGAGAGATTTATCTGTATATACACATAACTGAAAACCAAGTTTAAG 3368

Db 419 GAAGCTCAACTCAGAGAGAGATTTATCTGTATATACACATAACTGAAAACCAAGTTTAAG 478

QY 3369 CCCACCAATGCATGCTGATGCATGCCATATATTAATTAATGGGTAACTTTTATTTTATGA 3428

Db 479 CCCACCAATGCATGCTGATGCATGCCATATATTAATTAATGGGTAACTTTTATTTATGA 538

QY 3429 TCTCTACATAACAAGTGTGTTTGAAGGACATGTGAGCATATGATATGATCAATTT 3488

Db 539 TGTCTACATAACAAGTGTGATT--GGAAGGACATGTGAGCATATGATATGATCAATTT 596

QY 3489 TATGTTTTTTCTTTGTTTATATTTTGGGGAAGTTTAAATTTTAAAGGTAAA 3542

|||||

Db 597 TATGTTTTTTCTTT--GTTTATATTTTGGGGAAGTTTAAACATTTTAAAGGTATA 648

RESULT 8

B1227132 762 bp mRNA linear EST 11-JUL-2001

LOCUS 602952315F1 NIH_MGC_8 Homo sapiens cDNA clone IMAGE:5096643 5', mRNA sequence.

DEFINITION B1227132

ACCESSION B1227132.1 GI:14680576

VERSION B1227132

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 762)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L16M1868 row: n column: 04
High quality sequence stop: 659.
Location/Qualifiers

FEATURES

1. 762

Source

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5096643"

/clone_lib="NIH_MGC_8"

/tissue_type="Burkitt lymphoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dr priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit

BASE COUNT 225 a 176 c 178 g 183 t (Stratagene) and Superscript II RT (Life Technologies)."

Query Match 15.6%; Score 553.2; DB 13; Length 762;
Best Local Similarity 96.2%; Pred. No. 3.le-101;
Matches 612; Conservative 0; Mismatches 13; Indels 11; Gaps 4;
QY 890 CATTTCAGCAACAAACAGGCTTCATTCAGCTGCCAGGTGTGCCAGAAACACCTATTCTG 949
DB 2 CATTTCAGCAACAAACAGGCTTCATTCAGCTGCCAGGTGTGCCAGAAACACCTATTCTG 61
QY 950 AGAAAGAGCCAAAGAAATGATATAGGTGTAAAGAGGACTCTCAATTTTC-----AGGAT 1003
DB 62 AGAAAGAGCCAAAGAAATGATATAGGTGTAAAGAGGACTCTCAATTTTCAGAGGAAGGAT 121
QY 1004 CCAGTGTAGTGTACAGAGCCCTCTCTGTACACAAAGAGACTATTTCCAGATCCATATCTC 1063
DB 122 CCAGTGTAGTGTACAGAGCCCTCTCTGTACACAAAGAGACTATTTCCAGATCCATATCTC 181
QY 1064 CATGTGATGAAGCAAGCAAGACACAGATATCTACAAGTGGATAGAGCCCAAAATCTGCC 1123
DB 182 CATGTGATGAAGCAAGCAAGACACAGATATCTACAAGTGGATAGAGCCCAAAATCTGCC 241
QY 1124 GGGAGGATCTACAGATGCTATTAGATTGCCCTCTCTGGAGAGAGAGGATTGTCCGC 1183
DB 242 GGGAGGATCTACAGATGCTATTAGATTGCCCTCTCTGGAGAGAGAGGATTGTCCGC 301
QY 1184 CTGCAACCTGGATTTATACATGATCATCTCTCTGGCATCCCTGCTCTCTGGAA 1243
DB 302 CTGCAACCTGGATTTATACATGATCATCTCTCTGGCATCCCTGCTCTCTGGAA 361
QY 1244 CATTTCAGATGGAACCAAGAAATGTAGACCATGTCCAGCAGCAAGGAGCTGCACCTG 1303
DB 362 CATTTCAGATGGAACCAAGAAATGTAGACCATGTCCAGCAGCAAGGAGCTGCACCTG 421
QY 1304 GCTT-TGAATATAATGTGGGATGTCTCTCTGCAACATGAAACTTCTCTGCTTCAAT 1362
DB 422 GCTTCTGATATAAATGTGGGAATGTCTCTCTGCAACATGAAACTTCTCTGCTTCAAT 481
QY 1363 GTTGGGAATTCAGAGTGGATGG-TATGATGTTGGGAGTGGCTGGAGATCATATCCA 1421
DB 482 GTCCGGAATTCAGAGTGGATGGCAATGAATGTTGGGAGTGGCTGGAGATCATATCCA 541
QY 1422 GAGTGGGCTGGAGTGTTCGACAATGATTACCTGATCTTAAACTTGCATATCCAGG--- 1478
DB 542 GAGTGGGCTGGAGTGTTCGACAATGATTACCTGATCTTAAACTTGCATATCCAGGCA 601
QY 1479 ATTTAAACCAACCAATCTATGACTGGAGCCACGGG 1514
DB 602 TCTAAACCGAGCAACATCTATGACTGGAGCCACCGG 637

RESULT 9
BM752466
LOCUS
DEFINITION
BM752466
K-EST0029003 S9SNU601 Homo sapiens CDNA clone S9SNU601-17-H12 5',
mRNA linear EST 04-MAR-2002
mRNA sequence.
ACCESSION
BM752466
VERSION
BM752466.1
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 527)
AUTHORS
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
TITLE
21C Frontier Korean EST Project 2001
JOURNAL
Unpublished (2002)
COMMENT
Contact: Kim YS
Genome Research Center

Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 17 row: H column: 12
High quality sequence stop: 527.
Location/Qualifiers
1. 527

FEATURES

Source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S9SNU601-17-H12"
/clone_lib="S9SNU601"
/sex="M"
/tissue_type="Ascites"
/cell_type="Epithelial"
/cell_line="SNU-601"
/lab_host="Top10F"
/note="Organ: Stomach; Vector: pME18-FL3; Site:1: XhoI;
Site:2: XhoI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tabacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including Sfil
site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized with Superscript II using Sfil
oligo-dT primer. After first strand synthesis, RNA was
degraded by NaOH treatment and cDNA was amplified by PCR
reaction. The PCR products were digested with Sfil and
cloned into DraIII- digested pME18S-FL3 vector. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F, by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."

BASE COUNT 156 a 120 c 116 g 135 t
ORIGIN

Query Match 14.8%; Score 527; DB 14; Length 527;
Best Local Similarity 100.0%; Pred. No. 6.5e-96;
Matches 527; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1461 AAACCTTGCATATCCAGGATTTAAACCCACCAACATCTATGACTGGAGCCACGGTTCTGA 1520
DB 1 AAACCTTGCATATCCAGGATTTAAACCCACCAACATCTATGACTGGAGCCACGGTTCTGA 60
QY 1521 ACTAGGAAGAATAACATTTGTCTTTGAGACCCTCTCTTTCAGCTGACTGTGTTTGTACTT 1580
DB 61 ACTAGGAAGAATAACATTTGTCTTTGAGACCCTCTCTTTCAGCTGACTGTGTTTGTACTT 120
QY 1581 CATGTGGATATTAATAGAAAAAGTACAAATGTGGTAGAATCGTGGGTGGAAACCAAGA 1640
DB 121 CATGTGGATATTAATAGAAAAAGTACAAATGTGGTAGAATCGTGGGTGGAAACCAAGA 180
QY 1641 AAACAAGCTTACACCCATATCATCTTCAAGAATGCAACTTTTACATTTTACATGGGCATT 1700
DB 181 AAACAAGCTTACACCCATATCATCTTCAAGAATGCAACTTTTACATTTTACATGGGCATT 240
QY 1701 CCAGAGAACTATCAGGGTCAAGATAATAGACGGTTCATCAATGACATGGTGAAGATTTA 1760
DB 241 CCAGAGAACTATCAGGGTCAAGATAATAGACGGTTCATCAATGACATGGTGAAGATTTA 300
QY 1761 TTCTATCAGAGCCACTAATGCAAGTTGATGGGTGGGCTCTCATCGCTGCCTGTGCCCT 1820
DB 301 TTCTATCAGAGCCACTAATGCAAGTTGATGGGTGGGCTCTCATCGCTGCCTGTGCCCT 360
QY 1821 CGGTTCTGAACAGTGGGTTTCATCGTGTCTCCCTGCCCTCCAGGCCACTACATTTGAGAA 1880
DB 361 CGGTTCTGAACAGTGGGTTTCATCGTGTCTCCCTGCCCTCCAGGCCACTACATTTGAGAA 420
QY 1881 AGAAACCAACCAAGTGAAGGAATGTCCACTGACACCTACCTGTCCATACATCAGTCTTA 1940
DB 421 AGAAACCAACCAAGTGAAGGAATGTCCACTGACACCTACCTGTCCATACATCAGTCTTA 480
QY 1941 TGGCAAAAGAGGCTTGTATTCCATCGGGGCTGGAGTAAACCAATC 1987

```

|||||
Db 481 TGGCAAGAGGCTGTATTCCATCGCGCCTGGGAGTAAACAATC 527
|||||

RESULT 10
LOCUS BB384282 667 bp mRNA linear EST 24-OCT-2001
DEFINITION BB384282 RIKEN full-length enriched, 0 day neonate cerebellum Mus
            musculus cDNA clone C23003C19 3', mRNA sequence.
ACCESSION BB384282
VERSION BB384282.2 GI:16408254
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 667)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,
        Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
        M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Ozaki
        Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
        D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
        Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
        Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
JOURNAL On Jul 13, 2000 this sequence version replaced gi:9104048.
COMMENT Contact: Yoshihide Hayashizaki
        Laboratory for Genome Exploration Research Group, RIKEN Genomic
        Sciences Center (GSC), Yokohama Institute
        The Institute of Physical and Chemical Research (RIKEN)
        1-7-22 Suoh-cho, Tsukuba-shi, Ibaraki, Japan
        Tel: 81-45-503-9222
        Fax: 81-45-503-9216
        Email: genome-res@gsc.riken.go.jp.
        URL: http://genome.gsc.riken.go.jp/
        Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
        M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
        Normalization and subtraction of cap-trapper-selected cDNAs to
        prepare full-length cDNA libraries for rapid discovery of new
        genes. Genome Res. 10 (10), 1617-1630 (2000)
        wagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
        Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura
        S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
        Hayashizaki,Y.
        RIKEN integrated sequence analysis (RISA) system--384-format
        sequencing pipeline with 384 multicapillary sequencer. Genome Res.
        10 (11), 1757-1771 (2000)
        Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
        Y. and Hayashizaki,Y.
        Computer-based methods for the mouse full-length cDNA
        encyclopedia: real-time sequence clustering for construction of a
        nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
        Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
        K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
        Hayashizaki,Y.
        Computational Analysis of Full-length Mouse cDNAs Compared with
        Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
        Please visit our web site (http://genome.gsc.riken.go.jp/) for
        further details.
        cDNA library was prepared and sequenced in Mouse Genome
        Encyclopedia Project of Genome Exploration Research Group in Riken
        Genomic Sciences Center and Genome Science Laboratory in RIKEN.
        Division of Experimental Animal Research in Riken contributed to
        prepare mouse tissues.
FEATURES             Location/Qualifiers
     source            1..667
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="C23003C19"
                     /clone_lib="RIKEN full-length enriched, 0 day neonate
                     cerebellum"
                     /tissue_type="cerebellum"

```

RESULT 11
BQ551364

```

/dev_stage="0 day neonate"
/lab_host="DH10B"
/note="Site_1: Sali; Site_2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'-
GAGAGAGAGAGGATCCAGAGCTCTTTTCTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 479.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5'- GAGAGAGAGATTCGAGTTAAATAATATCCCTCCCTCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I."
BASE COUNT 169 a 136 c 167 g 194 t 1 others
ORIGIN
Query Match 14.4% Score 513.4; DB 10; Length 667;
Best Local Similarity 86.1% Pred. No. 3.4e-93;
Matches 568; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
QY 2185 ACCAACAATATAACAGACTTTACAGTAAAGAAATAGTGGCAGGTCAGATGATTACACA 2244
Db 8 ACCAACAATATTTCAGACTNTACGGTAAAGGAAATGGTGACAGGTCAGATGATTACACA 67
QY 2245 AATTTGGTAGGGCAATTTGATGCCAGTCAACAATATTTCCTTCTGAAAGTAAAGGTTTC 2304
Db 68 AATTTGGTAGGGCAATTTGATGCCAGTCAACAATATTTCCTTCTGAAAGTAAAGGTTTC 127
QY 2305 CGAGCAGCTTATCATCAACAATCCATTCATTCTGGCAGATACATTCATAGGAGTCACAGTT 2364
Db 128 CGGGCAGCCCTGCTGCACAGTCTATCATCTTGGCAGATATGTTTATAGTCTCACTGTT 187
QY 2365 GAAACACATTCGAAATATTAATATAAGAAAGATATGTTCCAGTGTCCACACAGCCAA 2424
Db 188 GACACTGCTTGCAGAAAGTAAATATCAAGGAGGATATGTTCCAGTGTCCACAGCCAA 247
QY 2425 ATACAGATGTCGCTTTCTTTTATAAGTCTCTACAGCAACAACATCTTGTATTAATGCG 2484
Db 248 GTACAGATGTTCTATTTATAGTCTTCCACAGCCACACATCTTGTATTAATGGA 307
QY 2485 CGATCAACTGCTGTGAAATAGAGGTGTAATCTTACTTAATCTGGAGCAGAGTGAATTC 2544
Db 308 CGATCGACTGCTGTGAAGATGAGGTGTAATCCATCAGACCTGGTGCAGGTGTGATATCA 367
QY 2545 GTCCCGCAGCAAGTCCCGCAGGTCACCTGATGGTGTACGTTCTATTTCTGTGGGAG 2604
Db 368 GTCCCGCAGCAAGTCCCGCAGGTCACCTGATGGTGTACGTTCTATTTCTATGGGAG 427
QY 2605 AGTGTGAAGCTTGCCTCTGTGTAGGAGATGACTTCCATGAGATTCAGGAGGCTGC 2664
Db 428 AGTGAGAAAGCTTGCCTCTGTGCACAGCAACATGACTTCCATGAGATTCAGGAGGCTGC 487
QY 2665 AAGAGAGGATTTTCAGGAAACCTTGTATGTGTGAATGAACCTTAAATGTCATTAAGGA 2724
Db 488 AAGAGAGGCTTTCAGGAAATATTAATGTAAGTGAACCTTAAATGTCATTAAGGA 547
QY 2725 ATTCTTTTCCCTGAGAAAAGTTGGCAACCTGTGAACGGTGTGACTTTTGGCTGAAGGTG 2784
Db 548 ATTCTTTTCCCTGAGAAAAGTTGTCAACCTGCAACCTTGTGACTTTTGGCTGAAGGTG 607
QY 2785 GGAGCGGCTGTGGAGCTTTTACTGCCGTTTGTGGTGGCTGTGACCTGTACTTCTGG 2844
Db 608 GGAGCTGGTGTGGGCGCTTTCACAGCCGTTTGTGGTGGCTTTAATACATGCTACTTTGG 667

```

```

LOCUS      BQ551364                634 bp      mRNA      linear      EST 20-JUN-2002
DEFINITION H4008H01-5 NIA Mouse 7.4K cDNA Clone Set Mus musculus cDNA clone
ACCESSION  BQ551364
VERSION    BQ551364.1 GI:21452250
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus.
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 634)
AUTHORS   VanBuren,V., Piao,Y., Dudekula,D.B., Qian,Y., Carter,M.G., Martin
            ,P.R., Stagg,C.A., Bassey,U., Aiba,K., Hamatani,T., Kargul,G.J.,
            Luo,A.G. and Ko,M.S.H.
TITLE     Assembly, verification, and initial annotation of NIA 7.4K mouse
            cDNA clone set
JOURNAL    Unpublished (2002)
COMMENT    Other_ESTs: H4008H01-3
            Contact: Yong Qian
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 3000, Baltimore, MD 21224-6820, USA
            Email: cdna@igsun.grc.nia.nih.gov
            This clone set has been freely distributed to the community. Please
            visit http://igsun.grc.nia.nih.gov/cDNA/NIA\_7.4k.html for details.
            Plate: H4008 row: H column: 01
            Seq primer: -21M13 Reverse
            High quality sequence stop: 634
POLYA-No. 180 a 143 c 165 g 146 t

FEATURES   Location/Qualifiers
            1..634
             /organism="Mus musculus"
             /strain="C57BL/6"
             /db_xref="niaEST:H4008H01-5"
             /db_xref="taxon:10090"
             /clone="H4008H01"
             /clone_lib="NIA Mouse 7.4K cDNA Clone Set"
             /sex="mixed"
             /dev_stage="mixed"
             /lab_host="DH10B"
             /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
clone is among a rearranged set of 7,407 clones from more
than 20 cDNA libraries."
BASE COUNT 180 a 143 c 165 g 146 t
ORIGIN

Query Match      14.4%; Score 512.8; DB 14; Length 634;
Best Local Similarity 89.3%; Pred. NO. 4.5e-93;
Matches 566; Conservative 0; Mismatches 62; Indels 6; Gaps 1;

QY  920  GCACGTGTGTCCTCCAGAAACCTATTCTGACAAAGGAGCCCAAGANTCTATAAGTGTA 979
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   1    GCACATGTGTCTCCAGAAACCTATTCTGAGAGGGTGCTAAAGAATCTATTAGTGTA 60
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY  980  AAGACGACTCTCAATTTC-----AGGATCCAGTGAGTGATACAGAGCCCTCCCTCTGA 1033
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   61   AGGAGGACTCCCGAGTTTCAGAGGAAGGAGCCAGTGAGTGTGGATCCGCCACCTGCA 120
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY  1034  CCACAAAGACATATTTCCAGATCCATCTCCATGTGATGAAGAAGAAACACACAGATAA 1093
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   121  CTACCAAAGACTATTTCAGATCCATACCCCTGTGACGAAGAAGGAACACACAGATAA 180
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY  1094  TGTACAAGTGGATAGAGCCCAAAATCTCCGGAGGAGATCTCACATGCTATTAGATGCC 1153
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   181  TGTACAAGTGGATAGAGCCCAAAATCTCGAGGGAGGATCTCACATGCTATTAGACTGC 240
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY  1154  CCCCTTCTGGAGAGAAGAGGATGTGCCCTTGCAACCTTGATTTTATAACAATGGAT 1213
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   241  CCCCTTCTGGAGAGAAGAGGATGTGCTCTCTTGCAACCCAGGATCTATACAATGGAT 300
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY  1214  CATCTCTTTGCCATCCCTCTCTCTCTGGAACATTTTCAGATGGAACCAAGAAATGTAGAC 1273
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   301  CGTCTTCTTGCCATCCCTCTCTCTCGGGGACGCTTTTCAGATGGAACAAAGGAATGCAACT 360
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

QY  1274  CATGTCACGAGGAACGAGCCTGCACCTGGCTTTTCAATATAAATGGTGAATGTCCTTC 1333
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   361  CATGTCGGCGGACAGAGCAGCAGCATTGGGTTTGAATATAAATGGTGAATGTCCTTC 420
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY  1334  CTGCAACATGAAAACTTCCCTGCTTCAATGTTGGGAATTTCAAAGTCCGATGGAATGAATG 1393
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   421  CTGCCAACATGAAAACTTCCCTGCTTCAATGTTGGGAATTTCAAAGTCCGATGGAATGAACG 480
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY  1394  GTTGGAGGTGCTGGAGATCATATCCAGATGGGGCTGGAGGTTCTGACAAATGATTACC 1453
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   481  GTTGGAGGTGCTGGAGATCATATCCGAGTGGAGCTGGAGGCTCTGACAATGACTATC 540
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY  1454  TGATCTTAAACTTCATATCCAGGATTTAAACCAACCAACATCTATGACTGGAGCCACGG 1513
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   541  TCATCTTAAACTTCGACATCCAGGATTTAAACCAACCAACGCTATGACTGGAGCCACAG 600
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY  1514  GTTCTGAACCTAGGAAGAATAACATTTGCTTTGA 1547
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db   601  GTTCGGAACCTAGGAAGGATAACATTTGCTTTGA 634
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
BM771726
LOCUS      BM771726                515 bp      mRNA      linear      EST 04-MAR-2002
DEFINITION K-EST0055671 S7SNU719s1 Homo sapiens CDNA clone S7SNU719s1-19-F04
ACCESSION  BM771726
VERSION    BM771726.1 GI:19101341
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 515)
AUTHORS   Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
TITLE     21C Frontier Korean EST Project 2001
JOURNAL    Unpublished (2002)
COMMENT    Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 19 row: F column: 04
            High quality sequence stop: 515.
            Location/Qualifiers
            1..515
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="S7SNU719s1-19-F04"
             /clone_lib="S7SNU719s1"
             /sex="M"
             /tissue_type="Stomach"
             /cell_type="Epithelial"
             /cell_line="SNU-719"
             /lab_host="Top10P"
             /note="Organ: Stomach; Vector: pcNS; Site_1: EcoRI;
            Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
            bacterial alkaline phosphatase (BAP) and then decapped
            with tabacco acid pyrophosphatase (TAP). The decapped
            intact mRNA was ligated with DNA-RNA linker including EcoR
            I site by treatment of T4 RNA ligase and the first strand
            cDNA was synthesized from oligo dt-selected mRNA by
            priming with dt-tailed vector. The dt-tailed vector was
            adjusted to have about 60nt. The cDNA vector was
            circularized with E. coli DNA ligase after digestion of
            EcoRI which site is also included in vector. An RNA strand
            converted to a DNA strand by Okayama-Berg method. The
            obtained cDNA vectors were used for transformation of

```


QY	1513	GGTTCGAAC	TAGGAAGAA	TAAACATTTGCTCTTTGAGACCCCTCTGTTTCAGCTGACTGTGTT	1572
Db	235	GGTTCGAAC	TAGGAAGAA	TAAACATTTGCTCTTTGAGACCCCTCTGTTTCAGCTGACTGTGTT	294
QY	1573	TTGTACTT	TCATGTTGG	GATATTAATAGAAAAAGTACAAATGTGGTAGAATCGTGGGGTGGGA	1632
Db	295	TCGTACTT	TCATGTTGG	GATATTAATAGAAAAAGTACAAATGTGGTAGAATCGTGGGGTGGGA	354
QY	1633	ACCAAGAAAA	AAACAAAGCTTAC	ACCCATATCATCTTCAAGAATGCAACTTTTACATTACAA	1692
Db	355	ACCAAGAAAA	AAACAAAGCTTAC	ACCCATATCATCTTCAAGAATGCAACTTTTACATTACAA	414
QY	1693	TGGGCATT	CCAGAGAACTATC	AGGGTCAAGATAATAGACGGTTTCATCAATGACATCGTG	1752
Db	415	TGGGCATT	CCAGAGAACTATC	AGGGTCAAGATAATAGACGGTTTCATCAATGACATCGTG	474
QY	1753	AAGATTTAT	TCTATCACAGC	CACTAATGCGAGTGGGTGGCGTCTCATGCGCGTGCC	1812
Db	475	AAGATTTAT	TCTATCACAGC	CACTAATGCGAGTGGGTGGCGTCTCATGCGCGTGCC	534
QY	1813	TGTGCCCT	TCGGTCTGAAC	AGTCGGGTTCATCGTGTGTCCTCCCTCCAGGCCACTAC	1872
Db	535	TGTGCCCT	TCGGTCTGAAC	AGTCGGGTTCATCGTGTGTCCTCCCTCCAGGCCACTAC	594
QY	1873	ATTGAGAA	AGAAACCAAC	1890	
Db	595	ATTGAGAA	AGAAACCAAC	612	
RESULT	15				
LOCUS	BI731892				
DEFINITION	603353640F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5361249 5',	801 bp	mRNA	linear	EST 20-SEP-2001
ACCESSION	BI731892				
VERSION	BI731892.1				
KEYWORDS	EST.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
TITLE	1 (bases 1 to 801)				
JOURNAL	NTH-MGC http://mgc.nci.nih.gov/ .				
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)				
	Contact: Robert Strausberg, Ph.D.				
	Email: cgabbs-remail.nih.gov				
	Tissue Procurement: The Cepko Laboratory				
	cDNA Library Preparation: Life Technologies, Inc.				
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)				
	DNA Sequencing by: Incyte Genomics, Inc.				
	Clone distribution: MGC clone distribution information can be				
	found through the I.M.A.G.E. Consortium/LLNL at:				
	http://image.llnl.gov				
	Plate: L14M11918 row: 0 column: 10				
	High quality sequence stop: 797.				
FEATURES	Location/Qualifiers				
source	1..801				
	/organism="Mus musculus"				
	/db_xref="taxon:10090"				
	/clone="IMAGE:5361249"				
	/clone_lib="NIH_MGC_94"				
	/tissue_type="retina"				
	/lab_host="DH10B (phage-resistant)"				
	/note="Organ: eye; Vector: pCMV-SpMT6; Site_1: NotI;				
	Site_2: SalI; Cloned unidirectionally; oligo-dT primed.				
	Average insert size 3.3 kb. Library enriched for				
	full-length clones and constructed by Life Technologies.				
	Note: this is a NIH_MGC Library."				
BASE COUNT	241 a 161 c 197 g 202 t				
ORIGIN					

Best Local Similarity 85.7%; Pred. No. 2.4e-89;		Matches 630; Conservative 0; Mismatches 98; Indels 7; Gaps 7;	
Qy	2375	TGAAAAATATTAATAAAGAGATATGTTCCAGTTCCAAACAGCCAAATACCAGATG	2434
Db	10	TGCAAAACGTTAAATATCAAGGAGGATATGTGCCAGGT-CACCGAGCCAAAGTACCAGATG	68
Qy	2435	TGCATTCTTTTATAAGTCTTCTACAGCAACAACATCTTGTTATTAAATGGCCGATCAACTG	2494
Db	69	TTCATTCTTTTATAAGTCTTCCACAGCCACCACATC-TGTATTAAATGGACGATCGACTG	127
Qy	2495	CTGTGAAATGAGGTGTAATCCTACTAAATCTGGAGCAGGAGTCAATTTTCAGTCCCCAGCA	2554
Db	128	CTGTGAAGATGAGGTGTAATCCCATGAGACCTGGTGCAGGTGTGATATCAGTCCCCAGCA	187
Qy	2555	AGTGCCACAGAGGTACCT-GTGATGGGTGTAGGTTCTATTTCCTGTGGGAGAGTGCTGAA	2613
Db	188	AGTGCCACAGCTGGCACCTGGGTGTGGTTGTACCTTCTACTTTCTATGGGAGAGTGCAGAA	247
Qy	2614	GCCTGCCCTCTGTGTACGGAGCATGACTTCCATGAGATTGAGGGAGCCTGCAAGAGAGCA	2673
Db	248	GCCTGCCCTCTGTGCACAGAACATGACTTCCATGAGATCGAGGAGCCTGCAAGAGAGGG	307
Qy	2674	TTTCAGGAAACCTTGTATGTGTGTAATGAACCTAAATGGTGCATTAAGGAATTTCTTTTG	2733
Db	308	CTTCAGGAAATATTATATGTATGTAATGAACCTAAGTGGTGCATTAAGGAATTTCTTTTG	367
Qy	2734	CCTGAGAAAAAGTTGGCAACCTGTGAAACGGTTGACTTTTCGCTCAAGGTGGAGCCGGT	2793
Db	368	CCCGAAAGAGAGTTGTCAACCTGGCAACTGTTGAC-TTTGGCTGAAAGTGGGAGCTGGT	426
Qy	2794	GTGGAGCTTTTACTGCGGTTTCTGCTGGCTCTGACCTCTACTTCTGGAAAAAGAAAT	2853
Db	427	GTGGCGCGTTTCACAGCGG-TTTGTTGGTGGCTTTAACATGCTACTTTTGCAAAAAGAA	485
Qy	2854	CAAAACTGGAATACAAATATTTCCAAAGTTAGTAATGACGACTAACTCAAAAGAGTGTGAA	2913
Db	486	CAAAAGCTGGAATACAAATATTTCTAAATTAGTAATGACGACTAACTCAAAAGAGTGTGAA	545
Qy	2914	CTCCCGCTGCAGACAGTTGTGCTATCATGGAAGGAGAGATATGAAGAGGAAGTTGTA	2973
Db	546	CTCCAGCTGCAGACAGCTGTGCTATCATGGAAGGAGAGATATGAAGAGGATGTGGTA	605
Qy	2974	TATTCCAAATAACAGTCACTACTAGGAAAACTCAAACTTTTGGCAACCAAGGAAAAAGAA	3033
Db	606	TATTCCAAACAAACAGTCACTACTAGGAAAACTTAAGTCTTGGCCACAAAGCAAAAGAT	665
Qy	3034	GACCATTTTGAATCTGTTCAACTGAAACCTCAAGATCCCCCAATATATGAAGAGAC-AG	3092
Db	666	GACCA-CTTGAGTCTGTTGAGCTGAAATCCCTCAAGATGCCCGCTATATGAAGAGACAA	724
Qy	3093	TGCTGTAGCCTTGAG	3107
Db	725	TGCTGTAGCATTCAG	739

Search completed: May 12, 2003, 07:48:31
Job time : 4363.66 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 00:12:42 ; Search time 4131.84 seconds
(without alignments)
13287.722 Million cell updates/sec

Title: US-10-073-333a-1
Perfect score: 3390
Sequence: 1 atgtgttcgcgcgcgggg.....ggtaaaaaaaaaaaaaa 3390

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_othr:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	660.4	19.5	697	14	BM723980
2	654.4	19.3	656	14	BM783507
3	624.8	18.4	930	12	BF035563
4	610.4	18.0	646	10	AW954806
5	553.2	16.3	762	13	BI227132
6	534.6	15.8	705	14	BQ182610

7	527	15.5	527	14	BM752466
8	513.4	15.1	667	10	BB384282
9	512.8	15.1	634	14	BO551364
10	507.4	15.0	515	14	BM771726
11	498.4	14.7	625	12	EG724280
12	493	14.5	541	10	BE237235
13	492.8	14.5	713	14	WQ772508
c 14	482.2	14.2	557	14	W60843
c 15	482.2	14.2	576	14	W60844
c 16	481.4	14.2	484	9	AA010992
c 17	478	14.1	478	10	AW65873
c 18	477	14.1	477	9	AI885785
c 19	472.6	13.9	585	14	BM974911
20	463	13.7	896	12	BG116864
21	449.4	13.3	508	13	BM538241
22	446	13.2	458	12	BF740039
23	435.6	12.8	1109	14	EQ896007
24	424.8	12.5	600	10	BB617824
25	421.8	12.4	656	10	BB626387
26	419.2	12.4	653	10	BB650824
c 27	414.4	12.2	507	9	AI797353
c 28	404	11.9	420	13	BI480715
c 29	397	11.7	445	9	AA010869
30	394.4	11.6	641	10	BB629056
31	391.8	11.6	667	13	BM491171
32	388.6	11.5	640	10	BB622295
33	381.8	11.3	859	14	BO730700
34	374.6	11.1	449	14	HI8553
35	374	11.0	374	9	AA644099
36	372	11.0	1773	11	AK017241
37	370.8	10.9	446	9	AI317059
c 38	369.4	10.9	395	10	AA444960
39	344.4	10.2	553	10	BB644517
40	332.6	9.8	631	13	BJ069789
41	331.6	9.8	387	10	BE081050
c 42	328.4	9.7	433	9	AI982900
c 43	325.2	9.6	801	13	BI731892
44	322.4	9.5	695	10	BB651047
45	317.2	9.4	689	9	AL644075

ALIGNMENTS

RESULT 1
BM723980
LOCUS
DEFINITION
UI-E-EQ1-a1y-a-05-0-UI.r1 UI-E-EQ1 Homo sapiens cDNA clone
UI-E-EQ1-a1y-a-05-0-UI 5', mRNA sequence.
ACCESSION
BM723980
VERSION
BM723980.1
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
TITLE
1 (bases 1 to 697)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL
MEDLINE
COMMENT
Genome Res. 6 (9), 791-806 (1996)
9704477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

697 bp
linear
EST 01-MAR-2002

Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).
Seq primer: M13 Reverse.

FEATURES

source

Location/Qualifiers

1. .697
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-E01-aly-a-05-0-UI"
/clone_lib="UI-E-E01"
/tissue_type="fetal eye"
/dev_stage="fetal"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: p773-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-E-E01 is a normalized cDNA library containing the following tissue(s): fetal eye. The library was constructed according to Bonaudo, Lennon and Soares, Genome Research, 6:731-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GCGGTATACC. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."
222 a 150 c 138 g 186 t 1 others
BASE COUNT
ORIGIN

Query Match 19.5%; Score 660.4; DB 14; Length 697;

Best Local Similarity 99.4%; Pred. No. 2.4e-12;

Matches 694; Conservative 0; Mismatches 12; Indels 3; Gaps 3;

QY 1835 CGGGTTACGTCGTCCCTCCAGGCCACTACAT-TGAGAAACAAACCAACAG 1893

DB 1 CGGGTTACGTCGTCCCTCCAGGCCACTACATNTGAGAAACAAACCAACAG 60

QY 1894 TCGAAGGAATGCCACCTGACACCTACCTGTCATACATCATGCTCTATGGCAAGAGGCT 1953

DB 61 TCGAAGGAATGCCACCTGACACCTACCTGTCATACATCATGCTCTATGGCAAGAGGCT 120

QY 1954 TGTATTCCATCGGGCTGGAGTAAACAATCAGGACCATCGGTTGCTATAGTGAC 2013

DB 121 TGTATTCCATCGGGCTGGAGTAAACAATCAGGACCATCGGTTGCTATAGTGAC 180

QY 2014 TGCTTTTCTACCATGAAAAGAAATCAGATTTTGCATATGACTTTAGCAACCTCAGC 2073

DB 181 TGCTTTTCTACCATGAAAAGAAATCAGATTTTGCATATGACTTTAGCAACCTCAGC 240

QY 2074 AGTGTGGCTCATTAATGAATGGCCCGCTTCACCTCCAAAGGAACAATACTTTCCAT 2133

DB 241 AGTGTGGCTCATTAATGAATGGCCCGCTTCACCTCCAAAGGAACAATACTTTCCAT 300

QY 2134 TTCTTCAATATCAGTTATGTGGCATGAGGGAAGAAGATGCTCTCTGTACCAACAT 2193

DB 301 TTCTTCAATATCAGTTATGTGGCATGAGGGAAGAAGATGCTCTCTGTACCAACAT 360

QY 2194 ATAACAGACTTTACAGTAAAGAAATAGTGGCAGGCTCAGATGATTACACAAA-TTTTGT 2252

DB 361 ATAACAGACTTTACAGTAAAGAAATAGTGGCAGGCTCAGATGATTACACAAA-TTTTGT 420

QY 2253 AGGGGCAATTTGATGCCAGTCAACAATATTCTTCTGAAAGTAAGGGTTTCCGAGCAGC 2312

DB 421 AGGGGCAATTTGATGCCAGTCAACAATATTCTTCTGAAAGTAAGGGTTTCCGAGCAGC 480

QY 2313 CTTATCATCAATCCATCATTTCTGGCAGATACATTCATAGAGTCACAGTTGAACACAC 2372

DB 481 CTTATCATCAATCCATCATTTCTGGCAGATACATTCATAGAGTCACAGTTGAACACAC 540

QY 2373 ATTGAAAAATATTATATATAAAGAGATATGTTCCAGTTTCCAAACAGCAATACACAGA 2432

|||||

DB 541 ATTGAAAAATATTATATAAAGAAGATATGTTCCAGTTTCCAAACCAATACACAGA 600

QY 2433 TGTGCATTTCTTTTATAAGTCTTTACAGCAACAACATCTTGTATTATATGCGCGATCAAC 2492

DB 601 TGTGCATTTCTTTTATAAGTCTTTACAGCAA-AACATCTTGTATTATATGCGCGATCAAC 659

QY 2493 TCCCTGTAAGATGAGGTCTAATCCTTACTAAATCTCGAG 2530

DB 660 TCCCTGTAAGATGAGGTCTAATCCTTACTAAATCTCGAG 697

RESULT 2

BM783507

LOCUS K-EST0061445 S7SNU719 Homo sapiens cDNA clone S7SNU719-36-H08 5', 656 bp mRNA linear EST 05-MAR-2002

DEFINITION mRNA sequence.

ACCESSION BM783507

VERSION BM783507.1 GI:19131739

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 656)

AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R., Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and Kim,Y.S.

TITLE 21C Frontier Korean EST Project 2001

JOURNAL Unpublished (2002)

COMMENT Contact: Kim YS

Genome Research Center

Korea Research Institute of Bioscience & Biotechnology

52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea

Tel: +82-42-860-4470

Fax: +82-42-860-4409

Email: yongsungemail.kr@kribb.re.kr

Plate: 36 row: H column: 08

High quality sequence stop: 656.

Location/Qualifiers

1. .656

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="S7SNU719-36-H08"

/clone_lib="S7SNU719"

/sex="M"

/tissue_type="Stomach"

/cell_type="Epithelial"

/lab_host="SNU-719"

/note="Organ: Stomach; Vector: pcNS; Site_1: EcoRI;

Site_2: NotI; The poly (A)+ RNA was dephosphorylated with

bacterial alkaline phosphatase (BAP) and then decapped

with tabacco acid pyrophosphatase (TAP). The decapped

intact mRNA was ligated with DNA-RNA linker including EcoR

I site by treatment of T4 RNA ligase and the first strand

cDNA was synthesized from oligo dT-selected mRNA by

priming with dT-tailed vector. The dT-tailed vector was

adjusted to have about 60nt. The cDNA vector was

circularized with E. coli DNA ligase after digestion of

EcoRI which site is also included in vector. An RNA strand

converted to a DNA strand by Okayama-Berg method. The

obtained cDNA vectors were used for transfection method.

The cDNA libraries constructed by this method are

full-length enriched cDNA library."

BASE COUNT 191 a 147 c 145 g 173 t

ORIGIN

Query Match 19.3%; Score 654.4; DB 14; Length 656;

Best Local Similarity 99.8%; Pred. No. 3.9e-12;

Matches 655; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1710 TAATCAGGCTCAAGATAATAGAGGGTTTCATCATGACATGGTGAAGATTTTATCTATAC 1769

Db	1	TAATCAGGGTCAAGATAATAGACGGTTTCATCAATGACATGGTGAAGATTATTCATATCAC	60
QY	1770	AGCCACTAATGAGTGGTGGCGCTCCCTCATGCGCTGCTGCGCTCGGTTCTGA	1829
Db	61	AGCCACTAATGAGTGGTGGCGCTCCCTCATGCGCTGCTGCGCTCGGTTCTGA	120
QY	1830	ACAGTCGGGTTCATGCTGTGTCCTCCCTCCAGGCCACTACATTGAGAAAGAACCAA	1889
Db	121	ACAGTCGGGTTCATGCTGTGTCCTCCCTCCAGGCCACTACATTGAGAAAGAACCAA	180
QY	1890	CCAGTCGAAGGAATGTCACCTGCACACCTACCTGTCCATACATCAGGTCTATGGCAAGA	1949
Db	181	CCAGTCGAAGGAATGTCACCTGCACACCTACCTGTCCATACATCAGGTCTATGGCAAGA	240
QY	1950	GGCTTGATTCATCGGGCTGGAGTGAAGAAACAAATCAGGACCATTCGGTTGCTATAG	2009
Db	241	GGCTTGATTCATCGGGCTGGAGTGAAGAAACAAATCAGGACCATTCGGTTGCTATAG	300
QY	2010	TGACTGCTTTTTTACCATGAAAGAAATCAGATTTTGCACATATGACTTTAGCAACCT	2069
Db	301	TGACTGCTTTTTTACCATGAAAGAAATCAGATTTTGCACATATGACTTTAGCAACCT	360
QY	2070	CAGCAGTGGGCTCATTAATGAATGGCCCCAGCTTCACCTCCAAAGGAACAAAATACTT	2129
Db	361	CAGCAGTGGGCTCATTAATGAATGGCCCCAGCTTCACCTCCAAAGGAACAAAATACTT	420
QY	2130	CCATTCTTCATATCAGTTTATGTGGCATCAGGGGAAGAGATGGCTCTGTACCAA	2189
Db	421	CCATTCTTCATATCAGTTTATGTGGCATCAGGGGAAGAGATGGCTCTGTACCAA	480
QY	2190	CAATATACAGACTTTACAGTAAAGAAATAGTGGCAGGTGAGATGATTACACAAATTT	2249
Db	481	CAATATACAGACTTTACAGTAAAGAAATAGTGGCAGGTGAGATGATTACACAAATTT	540
QY	2250	GTAGGGGATTGTATGCCAGTCAACAATATTCTTCTGAAAGTAAGGTTTCCGAGC	2309
Db	541	GTAGGGGATTGTATGCCAGTCAACAATATTCTTCTGAAAGTAAGGTTTCCGAGC	600
QY	2310	AGCCTTATCATCACATCCATCATCTGCGCAGATCATCATGAGGAGTGCAGTTG	2365
Db	601	AGCCTTATCATCACATCCATCATCTGCGCAGATCATCATGAGGAGTGCAGTTG	656
RESULT 3			
BF035563			
LOCUS	601457452F1 NIH_MGC_66	930 bp	mRNA linear EST 20-OCT-2000
DEFINITION	mRNA sequence.		
ACCESSION	BF035563		
VERSION	BF035563.1	GI:10743303	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	NIH-MGC http://mgi.nci.nih.gov/		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapbs-r@mail.nih.gov		
	Tissue Procurement: DCTD/DTF		
	cDNA Library Preparation: Life Technologies, Inc.		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	http://image.llnl.gov		
	plate: LLAM9597 row: g column: 13		
	High quality sequence stop: 647.		
FEATURES	Location/Qualifiers		
1..930			

Query Match 18.4%; Score 624.8; DB 12; Length 930;

Best Local Similarity 91.4%; Pred. No. 3.1e-115;

Matches 684; Conservative 0; Mismatches 62; Indels 2; Gaps 2;

QY 1661 TCATCTTCAAGAAATGCAACTTTTACATTTACATGGGCATTCAGAGAACTAAATCAGGGTC 1720

Db 1 TCATCTTCAAGAAATGCAACTTTTACATTTACATGGGCATTCAGAGAACTAAATCAGGGTC 60

QY 1721 AAGATAAATAGAGGTTTCATCAATGACATGGTGAAGATTTATTCATCAGACCCTAAATG 1780

Db 61 AAGATAAATAGAGGTTTCATCAATGACATGGTGAAGATTTATTCATCAGACCCTAAATG 120

QY 1781 CAGTTGATGGGTGGCTCTCATGCCGTGCCTGTCGCCCTCGGTTCTGAACAGTCGGGTT 1840

Db 121 CAGTTGATGGGTGGGTGCTCTCATGCCGTGCCTGTCGCCCTCGGTTCTGAACAGTCGGGTT 180

QY 1841 CATCGTGTGTCGCCCTCCAGGCGCACTACATTCAGAAAGAAACCAACCACTGCAAGG 1900

Db 181 CATCGTGTGTCGCCCTCCAGGCGCACTACATTCAGAAAGAAACCAACCACTGCAAGG 240

QY 1901 AATGTCACCTGACACCTACCTGTCATACATCAGTCTATGCGCAAGAGGCTTGATTTC 1960

Db 241 AATGTCACCTGACACCTACCTGTCATACATCAGTCTATGCGCAAGAGGCTTGATTTC 300

QY 1961 CATGGGGCTGGGAGTAAACAAATCAGGACCAATTCGGTTTCTGCTATAGTACTGCTTTT 2020

Db 301 CATGGGGCTGGGAGTAAACAAATCAGGACCAATTCGGTTTCTGCTATAGTACTGCTTTT 360

QY 2021 TCTACCATGAAAGAAATCAGATTTTGCACCTATGACCTTTAGCAACCTCAGCAGTGTGG 2080

Db 361 TCTACCATGAAAGAAATCAGATTTTGCACCTATGACCTTTAGCAACCTCAGCAGTGTGG 420

QY 2081 GCTCATTATGATGGCCCGCCAGCTTCACCTCCAAAGGACAAATACTTCCATTCTTCA 2140

Db 421 GCTCATTATGATGGCCCGCCAGCTTCACCTCCAAAGGACAAATACTTCCATTCTTCA 480

QY 2141 ATATCAGTTTATGTTGGGCATGAGGGGAAGAGATGCTCTCTGTACCAACAATATAACAG 2200

Db 481 ATATCAGTTTATGTTGGGCATGAGGGGAAGAGATGCTCTCTGTACCAACAATATAACAG 540

QY 2201 ACTTTACAGTAAAGAAATAGTGGCAGGTGAGATGATTACACAAATTTGGTAGGGGCAT 2260

Db 541 ACTTTACAGTAAAGAAATAGTGGCAGGTGAGATGATTACACAAATTTGGTAGGGGCAT 599

QY 2261 TTGATGCGCAGTCAACAATTTATTCCTTCTGAAAGTAAGGTTTCCGAGCAGCCTTATCAT 2320

Db 600 GTATGCCAGTCAACAATTTATTCCTTCTGAAAGTAAGGTTTCCGAGCAGCCTTATCAT 659

QY 2321 CACATATCCA-TCATTTCTGCGAGATCATTTAGGAGTGCAGATTTGAAACCAATGAA 2379

Db 660 CACATATCATTTCTTCTGGCGATACATTTCTAGGAGTGCAGAGTTGAAACCAATGGA 719

QY 2380 AATATTAAATAAAGAAAGATATGTTCC 2407

Db 720 AACTTTTATTTCCAGAGATATGTTCC 747

RESULT 4

AW954806

LOCUS

AW954806

646 bp

mRNA linear

EST 01-JUN-2000

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3861156"
/clone_lib="NIH_MGC_66"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pCMV-SpORT6; Site:1; Not1;
Site:2; SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

BASE COUNT 275 a 222 c 211 g 222 t

```
DEFINITION EST366876 MAGE resequences, MAGC Homo sapiens cDNA, mRNA sequence.
ACCESSION AW954806
VERSION AW954806.1 GI:8144489
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 646)
AUTHORS Hegde,P., Qi,R., Abernathy,K., Dhatap,S., Gaspard,R., Gay,C., Holt
,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
Quackenbush,J.
TITLE Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
JOURNAL Unpublished (2000)
COMMENT Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr... Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 69
Seq primer: Reverse.
FEATURES             Location/Qualifiers
     source
       1..646
         /organism="Homo sapiens"
         /db_xref="taxon:9606"
         /clone_lib="MAGE resequences, MAGC"
         /note="Vector: pBluescriptSKm"
BASE COUNT  189 a  143 c  142 g  171 t  1 others
ORIGIN
Query Match      18.0%; Score 610.4; DB 10; Length 646;
Best Local Similarity 98.8%; Pred. No. 2.7e-112;
Matches 636; Conservative 0; Mismatches 6; Indels 2; Gaps 2;
QY  1440 TGACAATGATTACCTGATCTTAAACTTCGATATCCAGAGATTAAACACCAACATCTAT 1499
DB   1  TGACAATGATTACCTGATCTTAAACTTCGATATCCAGAGATTAAACACCAACATCTAT 60
QY  1500 GACTGGAGCCAGGGTCTGAACTAGGAAGATAACATTTGCTTTTGAGACCTCTGTTC 1559
DB   61  GACTGGAGCCAGGGTCTGAACTAGGAAGATAACATTTGCTTTTGAGACCTCTGTTC 120
QY  1560 AGCTGACTGTGTTTGTACTTCATGCTGGTGGATATTAAAGAAAAGTACAAATGTGGTAGA 1619
DB   121 AGCTGACTGTGTTTGTACTTCATGCTGGTGGATATTAAAGAAAAGTACAAATGTGGTAGA 180
QY  1620 ATCGTGGGTGGAACCAAGAAAAGCAAGCTTACACCCATATCATCTTCAAGAAATGCCAAC 1679
DB   181 ATCGTGGGTGGAACCAAGAAAAGCAAGCTTACACCCATATCATCTTCAAGAAATGCCAAC 240
QY  1680 TTTTACATTTACATGGCATTCAGAGAACTAATCAGGGTCAAGATAATAGACGGTTTCAT 1739
DB   241 TTTTACATTTACATGGCATTCAGAGAACTAATCAGGGTCAAGATAATAGACGGTTTCAT 300
QY  1740 CAATGACATGGTGAAGATTTATTCTATCACGCCACTAATGCAAGTGTGATGGGGTGGCGTC 1799
DB   301 CAATGACATGGTGAAGATTTATTCTATCACGCCACTAATGCAAGTGTGATGGGGTGGCGTC 360
QY  1800 CTCATGCCGTGCTGTGCCCTCGGTTCTGAACAGTCGGGTTTCATCGTGTGTCCTGCCGCC 1859
DB   361 CTCATGCCGTGCTGTGCCCTCGGTTCTGAACAGTCGGGTTTCATCGTGTGTCCTGCCGCC 420
QY  1860 TCCAGGCCACTACATTGAGAAAAGAACCAACCAAGTGCAGAGGAATGCCACCTGACACCTTA 1919
DB   421 TCCAGGCCACTACATTGAGAAAAGAACCAACCAAGTGCAGAGGAATGCCACCTGACACCTTA 480
QY  1920 CTTGTCCATACATCAGGTCTATGGCAAGAGGCTTGTATTTCATCGGGGCCCTGGGAGTAA 1979
DB   481 CTTGTCCATACATCAGGTCTATGGCAAGAGGCTTGTATTTCATCGGGGCCCTGGGAGTAA 540
QY  1980 AACAATCAGGACCATTCCGGTTTGTATAGTACTGCTCTTTTCTACCATG-AAAAAGAA 2038
```

```
Db  541 ANACAATCAGACCCCTTCGGTTTCTGTATAGTACTGCTTTTCTACCATGAAAAAGAAA 600
QY  2039 ATCAGATTTTGCACATATGAC-TTTAGCAACCTTCAGCAGTGTGGG 2081
Db  601 ATCAGAGTTTGCACATATGACNTTTGGCAACCTTAACAGGGTGGG 644

RESULT 5
LOCUS BI227132
DEFINITION 602952315F1 NIH_MGC_8 Homo sapiens cDNA clone IMAGE:5096643 5',
mRNA sequence.
ACCESSION BI227132
VERSION BI227132.1 GI:14680576
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 762)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNL at:
http://image.llnl.gov
Plate: LLCM1868 row: n column: 04
High quality sequence stop: 659.
FEATURES             Location/Qualifiers
     source
       1..762
         /organism="Homo sapiens"
         /db_xref="taxon:9606"
         /clone_lib="IMAGE:5096643"
         /clone_lib="NIH_MGC_8"
         /tissue_type="Burkitt lymphoma"
         /lab_host="DH10B (phage-resistant)"
         /note="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2:
         EcoRI; cDNA made by oligo-dr priming. Directionally
         cloned into EcoRI/XhoI sites using the following 5'
         adaptor: GGCAGCAG(G). Size-selected >500bp for average
         insert size 1.8kb. Library constructed by Ling Hong in
         the laboratory of Gerald M. Rubin (University of
         California, Berkeley) using ZAP-cDNA synthesis kit
         (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT  225 a  176 c  178 g  183 t
ORIGIN
Query Match      16.3%; Score 553.2; DB 13; Length 762;
Best Local Similarity 96.2%; Pred. No. 7.7e-101;
Matches 612; Conservative 0; Mismatches 13; Indels 11; Gaps 4;
QY  890 CATTTCAGCAACAAACAGGTTTCATTTCACCTGCCAGGTGTGTCCAGAAACACCTATTCTG 949
DB   2  CATTTCAGCAACAAACAGGTTTCATTTCACCTGCCAGGTGTGTCCAGAAACACCTATTCTG 61
QY  950 AGAAGAGGCCAAAGAAATGTATAAGGTGTAAAGACGACTCTCAATTTTC-----AGGAT 1003
DB   62 AGAAGAGGCCAAAGAAATGTATAAGGTGTAAAGACGACTCTCAATTTTCAGAGGAAGGAT 121
QY  1004 CCAGTCAGTGTACAGAGCGCCCTCCCTGTACCACAAAGACTATTTCCAGATCCATATCTC 1063
DB   122 CCAGTCAGTGTACAGAGCGCCCTCCCTGTACCACAAAGACTATTTCCAGATCCATATCTC 181
QY  1064 CATGTGATGAAGAGGAAAGACACAGATTAATGTACAAAGTGGATAGAGCCCAAAATCTGCC 1123
DB   182 CATGTGATGAAGAGGAAAGACACAGATTAATGTACAAAGTGGATAGAGCCCAAAATCTGCC 241
```

RESULT 6	BQ182610/c	705 bp	mrna	linear	EST 30-APR-2002
LOCUS	BQ182610				
DEFINITION	UI-H-EU0-azx-a-23-0-UI.s1 NCICGAP_Carl Homo sapiens cDNA clone				
ACCESSION	BQ182610				
VERSION	BQ182610.1	GI:20358160			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 705)				
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .				
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),				
	Tumor Gene Index				
JOURNAL	Unpublished (1997)				
COMMENT	Contact: Robert Strausberg, Ph.D.				
	Email: cgapbs-ref@mail.nih.gov				
	Tissue Procurement: Dr. Jose Mercuende				
	cDNA Library preparation: Dr. M. Bento Soares, University of Iowa				
	cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa				
	DNA Sequencing by: Dr. M. Bento Soares, University of Iowa				
	Clone Distribution: Clone distribution information can be found				
	through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov				
	The following repetitive elements were found in this cDNA				
	sequence: 1-40, >AT-richLow_complexity (matched complement)				
	Seq primer: M13 FORWARD				
	POLYA=Yes.				

FEATURES	source
Location/Qualifiers	
1. .705	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/clone="IMAGE: 5854750"	
/clone_lib="NCI_CGAP_Carl"	
/tissue_type="Osteoarthritic Cartilage"	
/dev_stage="Adult"	
/lab_host="DH10B (Life Technologies)"	
/note="Organ: Knee; Vector: pT7r3-Pac (Pharmacia) with a modified polylinker; Site.1: EcoR I; Site.2: Not I; NCI_CGAP_Carl is a cDNA library containing the following tissue(s): Osteoarthritic Cartilage The library was constructed according to Bonaldo, Lennon and Soares,	

AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R., Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and Kim,Y.S.

TITLE 21C Frontier Korean EST Project 2001

JOURNAL Unpublished (2002)

COMMENT Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
plate: 17 row: H column: 12
High quality sequence stop: 527.

FEATURES

Source 1.527
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S9SNUG01-17-H12"
/clone_lib="S9SNUG01"
/sex="M"
/tissue_type="Asciites"
/cell_type="Epithelial"
/cell_line="SNU-601"
/lab_host="Top10F"
/note="Organ: Stomach; Vector: pME18-FL3; Site_1: XhoI; Site_2: XhoI; The poly (A)+ RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then decapped with tabacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including SfiI site by treatment of T4 RNA ligase and the first strand cDNA was synthesized with Superscript II using SfiI oligo-dT primer. After first strand synthesis, RNA was degraded by NaOH treatment and cDNA was amplified by PCR reaction. The PCR products were digested with SfiI and cloned into DraIII- digested pME18S-FL3 vector. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10F by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."

BASE COUNT 156 a 120 c 116 g 135 t

Query Match 15.5%; Score 527; DB 14; Length 527;
Best Local Similarity 100.0%; Pred. No. 1.6e-95;
Matches 527; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1461 AAACCTGGCATATCCAGGATTTAAACCAACCAACATCTATGACTGGAGCCACGGTCTCGA 1520

Db 1 AAACCTGGCATATCCAGGATTTAAACCAACCAACATCTATGACTGGAGCCACGGTCTCGA 60

QY 1521 ACTAGGAAGATAACATTTGTCTTTTGAGACCCCTCTGTTGAGCTGACTGTGTTTGTACTTT 1580

Db 61 ACTAGGAAGATAACATTTGTCTTTTGAGACCCCTCTGTTGAGCTGACTGTGTTTGTACTTT 120

QY 1581 CATGTGGATATTATAGAAAAGTACAATGTGGTAGAATCGTGGGGTGGGAACCAAGA 1640

Db 121 CATGTGGATATTATAGAAAAGTACAATGTGGTAGAATCGTGGGGTGGGAACCAAGA 180

QY 1641 AAAACAAGCTTACACCCATATCATCTTCAAGAATGCAACTTTTACATTTTACATGGGCATT 1700

Db 181 AAAACANGCTTACACCCATATCATCTTCAAGAATGCAACTTTTACATTTTACATGGGCATT 240

QY 1701 CCAGAGAACTAATCAGGGTCAAGATAAATAGACGGTTTCATCAATGATCGTGGTGAAGATTTA 1760

Db 241 CCAGAGAACTAATCAGGGTCAAGATAAATAGACGGTTTCATCAATGATCGTGGTGAAGATTTA 300

QY 1761 TTCTATCAGACCACTAATGACAGTTGATGGGTGGGTCTCTCATGCCGTGCGCTGCCCT 1820

Db 301 TTCTATCAGACCACTAATGACAGTTGATGGGTGGGTCTCTCATGCCGTGCGCTGCCCT 360

QY 1821 CGGTTCTGAACAGTCGGGTTCATCGTGTGCCCTGCCCTCCAGGCCACTACATTCAGAA 1880

Db 361 CGGTTCTGAACAGTCGGGTTCATCGTGTGCCCTCCAGCCCACTACATTTGAGAA 420

OY 1881 AGAAACCAACCAAGTCCAAGGAATGTCCACCTGACACCTACCTGTCCATACATCAGGTCTTA 1940

Db 421 AGAAACCAACCAAGTCCAAGGAATGTCCACCTGACACCTACCTGTCCATACATCAGGTCTTA 480

QY 1941 TGGCAAGAGGCTTGTATTTCATCGGGCTGGGAGTAAAAACAATC 1987

Db 481 TGGCAAGAGGCTTGTATTTCATCGGGCTGGGAGTAAAAACAATC 527

RESULT 8

LOCUS BB384282

DEFINITION BB384282 RIKEN full-length enriched, 0 day neonate cerebellum Mus

ACCESSION BB384282

VERSION BB384282.2 GI:16408254

KEYWORDS EST

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 667)

AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

TITLE Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.,

JOURNAL , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda

COMMENT , M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,

Okazaki,Y., Okido,T., Salto,R., Sakai,C., Sakai,K., Sano,H., Sasaki

D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,

Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,

Muramatsu,M. and Hayashizaki,Y.

RIKEN Mouse ESTs (Arakawa,T., et al. 2001)

Unpublished (2001)

On Jul 13, 2000 this sequence version replaced gi:9104048.

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsc.riken.go.jp,

URL: http://genome.gsc.riken.go.jp/

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh

, M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

wagii,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,

Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura

, S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and

Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara

, Y. and Hayashizaki,Y.

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa

, K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and

Hayashizaki,Y.

Computational Analysis of Full-length Mouse cDNAs Compared with

Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Please visit our web site (http://genome.gsc.riken.go.jp/) for

further details.

cDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues.

Location/Qualifiers

QY	1813	TGTGCCCTCGGTTCTGAACAGTCGGGTTTCATCGTGTGTCCTCCCTCCAGGCCACTAC	1872
Db	535	TGTGCCCTCGGTTCTGAACAGTCGGGTTTCATCGTGTGTCCTCCCTCCAGGCCACTAC	594
QY	1873	ATTGAGAAAGAAACCAAC	1890
Db	595	ATTGAGAAAGAAACCAAC	612
RESULT 12			
LOCUS	BE237235		
DEFINITION	146538 MARC 4BOV Bos taurus cDNA 5', mRNA linear		EST 25-APR-2001
ACCESSION	BE237235		
VERSION	BE237235.1	GI:9022025	
KEYWORDS	EST.		
SOURCE	COM.		
ORGANISM	Bos taurus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;		
	Bovidae; Bovinae; Bos.		
	1 (bases 1 to 541)		
REFERENCE	Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,		
AUTHORS	Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett		
	, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-Mckown, C.G.,		
	Perlea, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and		
	Keele, J.W.		
TITLE	Sequence evaluation of four pooled-tissue normalized bovine cDNA		
	libraries and construction of a gene index for cattle		
JOURNAL	Genome Res. 11 (4), 626-630 (2001)		
MEDLINE	21180013		
COMMENT	Contact: Smith TPL		
	USDA, ARS, US Meat Animal Research Center		
	PO Box 166, Clay Center, NE 68933-0166, USA		
	Tel: 402 762 4366		
	Fax: 402 762 4390		
	Email: smith@email.marc.usda.gov		
	Single pass sequencing. Bases called and alt trimmed with phred		
	v0.980904.e. Vector identified by cross_match with the -minscore 1		
	and -minmatch 12 options.		
	PCR Primers		
	FORWARD: AGGAACACGCTATGACCAT		
	BACKWARD: GTTTTCCCAGTCACGACG		
	Plate: 46 row: K column: 21		
	Seq primer: ATTTAGGTGACACTATAG.		
FEATURES			
source	Location/Qualifiers		
	1..541		
	/organism="Bos taurus"		
	/db_xref="taxon:9913"		
	/clone_lib="MARC 4BOV"		
	/tissue_type="pooled"		
	/lab_host="DH10B"		
	/note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;		
	Library made from pooled tissue from day 20 and day 40		
	embryos."		
BASE COUNT	155 a 104 c 134 g 148 t		
ORIGIN			
Query Match	14.5%;	Score 493;	DB 10; Length 541;
Best Local Similarity	94.5%;	Pred. No. 1e-88;	
Matches	511; Conservative	0; Mismatches	30; Indels 0; Gaps
QY	1161	TGGAGAGAAGAGGATTTCGCGCTTGCAACCCCTGGATTTTATTAACAAATGGATCATCTTC	1220
Db	1	TGGAGAGAAGAGGATTTCGCGCTTGCAACCCCTGGATTTTATTAACAAATGGATCATCTTC	60
QY	1221	TGGCATCCCTGTCTCTCGGAAATTTTCAGATGGAACCAAGAAATGTAGACCATCTTC	1280
Db	61	CTGCCATCCCTGTCTCTCGGAAATTTTCAGATGGAACCAAGAAATGTAGATCATCTTC	120
QY	1281	AGCAGAACGAGCCTGCATCTGGCTTTGAATATATAATGGTGGAAATGTCTTCTCTGCAA	1340
Db	121	AGCAGAACGAGCCTGCATCTGGCTTTGAATATATAATGGTGGAAATGTCTTCTCTGCAA	180

REFERENCE 1 (bases 1 to 557)
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, F., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohldmann, P., and Wilson, R.
 TITLE The WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 1674 Std Error: 0.00
 Seq primer: mob.REGA+ET
 High quality sequence stop: 391.
 Location/Qualifiers
 1. 557
 /organism="Homo sapiens"
 /db_xref="GDB:1267178"
 /db_xref="taxon:9606"
 /clone_lib="IMAGE:341803"
 /sex="unknown"
 /dev_stage="19 weeks"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: heart; Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAACTCTGAAGTGGAGCGCGCATCTTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M.Fatima Bonaldo. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung NBHL19W."

BASE COUNT 180 a 117 c 107 g 152 t 1 others
 ORIGIN
 Query Match 14.2%; Score 482.2; DB 14; Length 557;
 Best Local Similarity 98.2%; Pred. No. 1.5e-86;
 Matches 550; Conservative 0; Mismatches 4; Indels 6; Gaps 6;

QY 1916 CCTACCTGTCCATACATCAGGCTTATGGCAAGAGGCTTGTTATTCATCGCGGCTGGGA 1975
 |||||
 Db 1 CCTACCTGTCCATACATCAGGCTTATGGCAAGAGGCTTGTTATTCATCGCGGCTGGGA 60
 |||||

QY 1976 GTAAACAAATCAGGACCATTCGGTTTGTATAGTACTGCTTTTCTACCATGAAAG 2035
 |||||
 Db 61 GTAAACAAATCAGGACCATTCGGTTTGTATAGTACTGCTTTTCTACCATGAAAG 120
 |||||

QY 2036 AAAATCAGATTTTGCACTATGACTTTAGCAACTCAGCAGTGGCTCATTAATGAATG 2095
 |||||
 Db 121 AAAATCAGATTTTGCACTATGACTTTAGCAACTCAGCAGTGGCTCATTAATGAATG 180
 |||||

QY 2096 GCCCAGCTTCCACCTCCAAAGGAACAAATACTTCCATTTCTCAATATCAGTTATGTG 2155
 |||||
 Db 181 GCCCAGCTTCCACCTCCAAAGGAACAAATACTTCCATTTCTCAATATCAGTTATGTG 240
 |||||

QY 2156 GGCATGAGGGGAGAGATGGCTCTCTGTACCAACAATATACAGACTTTACA-GTAA 2214
 |||||

Db 241 GGCATGAGGGGAGAGATGGCTCTCTGTACCAACAATATACAGACTTTACACGTTAAA 300
 |||||

QY 2215 GAAATAGTGGCAGGTCAGATGATTACAAATTTGGTAGGGGCAATTCATGCCAGTCA 2274
 |||||

Db 301 GAAATAGTGGCAGGTCAGATGATTACAAATTTGGTAGGGGCAATTCATGCCAGTCA 359
 |||||

QY 2275 ACAATTTATTCCTCTGAAAGTAA-GGGTTTCGAGCAGCGCTTTATCATCACAAAT-CCATCA 2332
 |||||

Db 360 ACAATTATTCCCTTCGAAAGTAAGGGTTCCGAGCAGCGCTTATCATCACAAATCCCATCA 419
 QY 2333 TTCTGGCAGATACATTCATAGCAGTCACAGTTGAACACACATTTGAAATATATATATAA 2392
 |||||
 Db 420 TTCTGGCAGATACATTCATAGCAGTCACAGTTGAACACACATTTGAAATATATAT-A 477
 |||||

QY 2393 AAGAAGATATGTTCCAGTTTCCAACAGCAATACCAGATGTCATTTCTTTTATAAGT 2452
 |||||
 Db 478 AAGAAGATATGTTCCAGTTTCCAACAGCAATACCAGATGTCATTTCTTTTATAAGT 537
 |||||

QY 2453 CTCTACAGCAACACATCT 2472
 |||||

Db 538 CTCTAAAGCAACAACATCT 557
 |||||

RESULT 15
 W60844/c
 LOCUS 576 bp mRNA linear EST 15-OCT-1996
 DEFINITION IMAGE:341803 3', mRNA sequence.
 ACCESSION W60844
 VERSION W60844.1 GI:1367602
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 576)
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, F., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohldmann, P., and Wilson, R.
 TITLE The WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 1674 Std Error: 0.00
 Seq primer: mob.REGA+ET
 High quality sequence stop: 336.
 Location/Qualifiers
 1. 576
 /organism="Homo sapiens"
 /db_xref="GDB:1267178"
 /db_xref="taxon:9606"
 /clone_lib="IMAGE:341803"
 /clone_lib="Soares_fetal_heart_NbHL19W"
 /sex="unknown"
 /dev_stage="19 weeks"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: heart; Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAACTCTGAAGTGGAGCGCGCATCTTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M.Fatima Bonaldo. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung NBHL19W."

BASE COUNT 171 a 112 c 108 g 182 t 3 others
 ORIGIN
 Query Match 14.2%; Score 482.2; DB 14; Length 576;
 Best Local Similarity 97.1%; Pred. No. 1.5e-86;
 Matches 501; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

QY 2860 AAAAGAAGACACCTTTTGAATCTGTT-CAACTGAAAACTCAAGATCCCAAAATATATGA 2918
||||| ||| ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| |||||||||
Db 516 AAAAGGAGACCCATTTTGAATCTGTTCCAACTGAACCTCAAGATCCCAAAATATATGA 457
QY 2919 AGAGACAGTGTCTAGCCCTTGAGACTTAATGAACAAGAAACCTGCTCTAGTTTACAGGA 2978
||||||| ||||||||| ||||||||| ||| ||||||||| ||||||||| ||||||||| |||||||||
Db 456 AGAGACAGTCTGTAGCCCTTGAGACTTAATGAACAAGAAACCTGCTCTAGTTTACAGGA 397
QY 2979 CCATATTTTAGGGTCTGTCTCATACCTGTACATTTGGTGATCTCACAGAGGAGGCCAT 3038
||||||| ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| |||||||||
Db 396 CCATATTTTAGGGTCTGTCTCATACCTGTACATTTGGTGATCTCACAGAGGAGGCCAT 337
QY 3039 GCCGCTGAAAAAGGAGGAGATTGAAACATTTGATTCCTTATCACATGGTCAAGTACCT 3098
||||||| ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| |||||||||
Db 336 GCCGCTGAAAAAGGAGGAGATTGAAACATTTGATTCCTTATCACATGGTCAAGTACCT 277
QY 3099 TGCCCAATAAAGGAAGCAAAATGATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAGGA 3158
||||||| ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| |||||||||
Db 276 TGCCCAATAAAGGAAGCAAAATGATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAGGA 217
QY 3159 AGAGATTTTATCTGTATATACACATAACTGAAACCAAGTTTAAAGCCACCAATGCACCTGC 3218
||||||| ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| |||||||||
Db 216 AGAGATTTTATCTGTATATACACATAACTGAAACCAAGTTTAAAGCCACCAATGCACCTGC 157
QY 3219 TGATGCATGCCATATAAATTAATGGGTAACTTTTATCTTTTATGATGTCTTACATAACAAGT 3278
||||||| ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| |||||||||
Db 156 TGATGCATGCCATATAAATTAATGGGTAACTTTTATCTTTTATGATGTCTTACATAACAAGT 97
QY 3279 GTGATTTGGAAGGCACATGTGAGCATATGCATATGATCCCAATTTATGTTTTCTTTGT 3338
||||||| ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| |||||||||
Db 96 GTGATTTGGAAGGCACATGTGAGCATATGCATATGATCCCAATTTATGTTTTCTTTGT 37
QY 3339 TTATATTTTGGGAAAAATTAATTTTAAAGGTA 3374
||||||| ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| |||||||||
Db 36 TTATATTTTGGGAAAAATTAATTTTAAAGGTA 1

Search completed: May 12, 2003, 07:48:02
Job time : 4163.34 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 05:40:13 ; Search time 521.238 Seconds
(without alignments)
8082.236 Million cell updates/sec

Title: US-10-073-333A-1
Perfect score: 3390
Sequence: 1 atgctgttcgcgcgcgggg.....ggtataaaaaaaaaaaaaa 3390

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 783854 seqs, 621352466 residues

Total number of hits satisfying chosen parameters: 1567708

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA.*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3390	100.0	3390	9	US-10-140-164-1
2	3214	94.8	3556	9	US-10-140-164-3
3	1499.6	44.2	1737	9	US-10-002-050-19
4	1499.6	44.2	1737	9	US-10-002-304-19
5	1499.6	44.2	1737	12	US-10-003-152-19
6	1182.6	34.9	1508	9	US-10-002-050-9
7	1182.6	34.9	1508	9	US-10-002-304-9
8	1182.6	34.9	1508	12	US-10-003-152-9
9	901	26.6	3501	9	US-10-028-072-37
10	901	26.6	3501	9	US-10-121-049-37
11	901	26.6	3501	9	US-10-123-904-37
12	901	26.6	3501	9	US-10-140-470-37
13	901	26.6	3501	9	US-10-175-746-37
14	901	26.6	3501	9	US-10-176-918-37
15	901	26.6	3501	9	US-10-176-921-37
16	901	26.6	3501	9	US-10-137-865-37
17	901	26.6	3501	9	US-10-140-474-37
18	901	26.6	3501	9	US-10-142-431-37
19	901	26.6	3501	9	US-10-143-114-37

20	901	26.6	3501	9	US-10-140-002-37	Sequence 37, Appl
21	901	26.6	3501	9	US-10-142-419-37	Sequence 37, Appl
22	901	26.6	3501	9	US-10-123-262-37	Sequence 37, Appl
23	901	26.6	3501	9	US-10-142-423-37	Sequence 37, Appl
24	901	26.6	3501	9	US-10-121-050-37	Sequence 37, Appl
25	901	26.6	3501	9	US-10-141-755-37	Sequence 37, Appl
26	901	26.6	3501	9	US-10-143-032-37	Sequence 37, Appl
27	901	26.6	3501	9	US-10-123-108-37	Sequence 37, Appl
28	901	26.6	3501	9	US-10-123-236-37	Sequence 37, Appl
29	901	26.6	3501	9	US-10-123-261-37	Sequence 37, Appl
30	901	26.6	3501	9	US-10-140-921-37	Sequence 37, Appl
31	901	26.6	3501	9	US-10-140-928-37	Sequence 37, Appl
32	901	26.6	3501	9	US-10-121-045-37	Sequence 37, Appl
33	901	26.6	3501	9	US-10-123-292-37	Sequence 37, Appl
34	901	26.6	3501	9	US-10-123-903-37	Sequence 37, Appl
35	901	26.6	3501	9	US-10-124-819-37	Sequence 37, Appl
36	901	26.6	3501	9	US-10-124-822-37	Sequence 37, Appl
37	901	26.6	3501	9	US-10-140-925-37	Sequence 37, Appl
38	901	26.6	3501	9	US-10-160-498-37	Sequence 37, Appl
39	901	26.6	3501	9	US-10-121-041-37	Sequence 37, Appl
40	901	26.6	3501	9	US-10-121-043-37	Sequence 37, Appl
41	901	26.6	3501	9	US-10-121-047-37	Sequence 37, Appl
42	901	26.6	3501	9	US-10-123-215-37	Sequence 37, Appl
43	901	26.6	3501	9	US-10-123-902-37	Sequence 37, Appl
44	901	26.6	3501	9	US-10-123-908-37	Sequence 37, Appl
45	901	26.6	3501	9	US-10-123-909-37	Sequence 37, Appl

ALIGNMENTS

RESULT 1
US-10-140-164-1
; Sequence 1, Application US/10140164
; Publication No. US20030072736A1
; GENERAL INFORMATION:
; APPLICANT: Baker et al.
; TITLE OF INVENTION: Human Tumor Necrosis Factor Receptor Tr16
; FILE REFERENCE: PF514C1
; CURRENT APPLICATION NUMBER: US/10/140,164
; CURRENT FILING DATE: 2002-05-08
; PRIOR APPLICATION NUMBER: 09/637,856
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/148,348
; PRIOR FILING DATE: 1999-08-12
; PRIOR APPLICATION NUMBER: 60/148,683
; PRIOR FILING DATE: 1999-08-13
; PRIOR APPLICATION NUMBER: 60/148,870
; PRIOR FILING DATE: 1999-08-13
; PRIOR APPLICATION NUMBER: 60/148,758
; PRIOR FILING DATE: 1999-08-16
; PRIOR APPLICATION NUMBER: 60/149,181
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: 60/149,453
; PRIOR FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: 60/149,498
; PRIOR FILING DATE: 1999-08-19
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 3390
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2892)
US-10-140-164-1

Query Match 100.0% Score 3390; DB 9; Length 3390;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3390; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCTGTTCCGCGCGCGCGCGGTACGGGCGCGGTGGGGCGCGCGGGAGGCT 60

||||| 1 ATGCTGTTCCGCGCGGCGGCGGTACGGGCGAGGGCTGGGGCGCGCGGAGGCT 60
QY 61 CCCGCGCGCGGCGCTCCGCGCGCTGGAGCGCGCGCTGGATTTGCTGCTGGCGGCTCGCC 120
Db 61 CCCGCGCGCGGCGCTCCGCGCGCTGGAGCGCGCGCTGGATTTGCTGCTGGCGGCTCGCC 120
QY 121 GCGTCCAGCGGCGCTGGGCTGGGACCTGCGCTCTCCAGCGCGCGCGCTTCCCTCT 180
Db 121 GCGTCCAGCGGCGCTGGGCTGGGACCTGCGCTCTCCAGCGCGCGCGCTTCCCTCT 180
QY 181 TCCAGGAGAAAGATTATCACTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGG 240
Db 181 TCCAGGAGAAAGATTATCACTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGG 240
QY 241 AGAGTTGCCATTCCAAATTTCTGACGTGACCTGCTGCGCTGCCCTGACCCAGTGAGAGGC 300
Db 241 AGAGTTGCCATTCCAAATTTCTGACGTGACCTGCTGCGCTGCCCTGACCCAGTGAGAGGC 300
QY 301 AAAGAAATGCACCTTCTCTGCTCTTCTGGAGAGTATCTAGAAATGAAGAACAGATATGC 360
Db 301 AAAGAAATGCACCTTCTCTGCTCTTCTGGAGAGTATCTAGAAATGAAGAACAGATATGC 360
QY 361 AGTAAGTGTGTGAAGGACCACTATTTCTTGGCAGTGGCATCAAAATTTGATGAATGGGAT 420
Db 361 AGTAAGTGTGTGAAGGACCACTATTTCTTGGCAGTGGCATCAAAATTTGATGAATGGGAT 420
QY 421 GAATTTCCGCGGAGGATTTCTAACATCGCAACATTTCAATGACACTGTGGCGGCGCTTCT 480
Db 421 GAATTTCCGCGGAGGATTTCTAACATCGCAACATTTCAATGACACTGTGGCGGCGCTTCT 480
QY 481 GACAGAGCCAGAGCGGTGTAAACACTCTTCTGGATCCCTCGTGGAACTACATAGAA 540
Db 481 GACAGAGCCAGAGCGGTGTAAACACTCTTCTGGATCCCTCGTGGAACTACATAGAA 540
QY 541 TCTAATCGTGTGACGCGGTGTCTTTGATCTATGCTGTCCACCTTAAAGATCAGGC 600
Db 541 TCTAATCGTGTGACGCGGTGTCTTTGATCTATGCTGTCCACCTTAAAGATCAGGC 600
QY 601 TATGCTCTTCTTGTAGTACCAGTATGCGACAACACATCTTCTTGTAGTCTTTATTCAA 660
Db 601 TATGCTCTTCTTGTAGTACCAGTATGCGACAACACATCTTCTTGTAGTCTTTATTCAA 660
QY 661 AATGATCAGTCCAGGAGATGGACACCACTGACAACTGAGGTGAACTTACAGACAAT 720
Db 661 AATGATCAGTCCAGGAGATGGACACCACTGACAACTGAGGTGAACTTACAGACAAT 720
QY 721 GGAGATGGGCTCTCATTTCTGTATGCTGAATCAGGCACAAACATCTCTACTGGAGA 780
Db 721 GGAGATGGGCTCTCATTTCTGTATGCTGAATCAGGCACAAACATCTCTACTGGAGA 780
QY 781 ACTACAGGATCTTATGGGTTCTAAGGCGGTCAAGCGCTGTGCTGTTAAAAATATCACA 840
Db 781 ACTACAGGATCTTATGGGTTCTAAGGCGGTCAAGCGCTGTGCTGTTAAAAATATCACA 840
QY 841 ATTGAAGGGGTGGCGTACACATCAGAAATGTTTTCCTTGAAGCCAGGACATTCAGCAAC 900
Db 841 ATTGAAGGGGTGGCGTACACATCAGAAATGTTTTCCTTGAAGCCAGGACATTCAGCAAC 900
QY 901 AAACCAAGGTTCTTCAACTGCCAGGTGTGTCCAGAAACACCTATTCTGAGAAAGAGGC 960
Db 901 AAACCAAGGTTCTTCAACTGCCAGGTGTGTCCAGAAACACCTATTCTGAGAAAGAGGC 960
QY 961 AAAGAATGTATAAGGTGTAAAGACGACTCTCAATTTTCCAGATCCAGTGTACAGAG 1020
Db 961 AAAGAATGTATAAGGTGTAAAGACGACTCTCAATTTTCCAGATCCAGTGTACAGAG 1020
QY 1021 CGCCCTCCCTGTATCCACAAAGACTATTTCCAGATCCACTACTGTGTATGAAGAGGA 1080
Db 1021 CGCCCTCCCTGTATCCACAAAGACTATTTCCAGATCCACTACTGTGTATGAAGAGGA 1080
QY 1081 AAGACACAGATATGTACAGTGGATAGAGCCCAAAATCTGCGGGAGGATCTCACAGAT 1140
|||||

Db 1081 AAGACACAGATATGTACAGTGGATAGAGCCCAAAATCTGCCGGAGGATCTCACAGAT 1140
QY 1141 GCTATTAGATTGCCCTTCTTGAGAGAGAAAGATTTGCTCGCCTTGCAACCCCTGGATTT 1200
Db 1141 GCTATTAGATTGCCCTTCTTGAGAGAGAAAGATTTGCTCGCCTTGCAACCCCTGGATTT 1200
QY 1201 TATACAAATGGATCATCTTCTTGCCATCCCTGCTCTGGAACAATTTTCAGATGGAACC 1260
Db 1201 TATACAAATGGATCATCTTCTTGCCATCCCTGCTCTGGAACAATTTTCAGATGGAACC 1260
QY 1261 AAAGAATGTAGACCATGTCCAGCAGGAACGGAGCTTGCACCTTGGCTTTGAAATATATATG 1320
Db 1261 AAAGAATGTAGACCATGTCCAGCAGGAACGGAGCTTGCACCTTGGCTTTGAAATATATATG 1320
QY 1321 TGGAAATGCTTCTTCCGGAACATGAAACTTCCGCTTCAATGTTGGAAATTCAAAGTGC 1380
Db 1321 TGGAAATGCTTCTTCCGGAACATGAAACTTCCGCTTCAATGTTGGAAATTCAAAGTGC 1380
QY 1381 GATGAATGAATGGTTGGAGGTGGCTGAGATCATATCCAGAGTGGGCTGGAGGTTCT 1440
Db 1381 GATGAATGAATGGTTGGAGGTGGCTGAGATCATATCCAGAGTGGGCTGGAGGTTCT 1440
QY 1441 GACAATGATTAACCTGATCTTAACTTGCATATCCAGGATTTAAACACCAACATCTATG 1500
Db 1441 GACAATGATTAACCTGATCTTAACTTGCATATCCAGGATTTAAACACCAACATCTATG 1500
QY 1501 ACTGAGGACCGGTTCTGAACTAGGAAGATTAACATTTGCTTTGAGACCTCTGTTCA 1560
Db 1501 ACTGAGGACCGGTTCTGAACTAGGAAGATTAACATTTGCTTTGAGACCTCTGTTCA 1560
QY 1561 GCTGACTGTGTTTTCTACTTTCATGTTGATATTAAGAAAAGTACAAATGTGGTAGAA 1620
Db 1561 GCTGACTGTGTTTTCTACTTTCATGTTGATATTAAGAAAAGTACAAATGTGGTAGAA 1620
QY 1621 TCGTGGGTGGAACCAAGAAACAAAGCTTACACCCATATCATCTTCAAGAAATGCAACT 1680
Db 1621 TCGTGGGTGGAACCAAGAAACAAAGCTTACACCCATATCATCTTCAAGAAATGCAACT 1680
QY 1681 TTTACATTTACATGGGCATTTCCAGAGAACTAATCAGGGTCAAGATATAGACGTTTCATC 1740
Db 1681 TTTACATTTACATGGGCATTTCCAGAGAACTAATCAGGGTCAAGATATAGACGTTTCATC 1740
QY 1741 AATGACATGTTGAAGATTTATTTCTATCAGAGCACTAATGCGAGTTGATGGGTGGCTCC 1800
Db 1741 AATGACATGTTGAAGATTTATTTCTATCAGAGCACTAATGCGAGTTGATGGGTGGCTCC 1800
QY 1801 TCAATGCGGTGCTGCTGCTCGCTTCTGAACAGTGGGTTCAATGCTGCTGCTGCTGCTGCT 1860
Db 1801 TCAATGCGGTGCTGCTGCTCGCTTCTGAACAGTGGGTTCAATGCTGCTGCTGCTGCTGCT 1860
QY 1861 CGAGGCACTACATTTGAGAAAGAAACCAACCACTGCGAGGATGTCACCTGACACCTAC 1920
Db 1861 CGAGGCACTACATTTGAGAAAGAAACCAACCACTGCGAGGATGTCACCTGACACCTAC 1920
QY 1921 CTGTCCATACATCAGGTCTATGGAAGAGGCTTGTATTTCCATGCGGCTGGAGTAA 1980
Db 1921 CTGTCCATACATCAGGTCTATGGAAGAGGCTTGTATTTCCATGCGGCTGGAGTAA 1980
QY 1981 AACAAATCAGGACCATTTGCTTATAGTACTGCTTTTCTTACCATGAAAGAAAT 2040
Db 1981 AACAAATCAGGACCATTTGCTTATAGTACTGCTTTTCTTACCATGAAAGAAAT 2040
QY 2041 CAGATTTTGCATATGACATTTACCAACCTCAGCAGTGGGCTCATTAATGAATGGCCC 2100
Db 2041 CAGATTTTGCATATGACATTTACCAACCTCAGCAGTGGGCTCATTAATGAATGGCCC 2100
QY 2101 AGCTTCCACTTCCAAAGAAACAAATACCTTCAATTTCTCAATATCAGTTTATGTTGGCAT 2160
Db 2101 AGCTTCCACTTCCAAAGAAACAAATACCTTCAATTTCTCAATATCAGTTTATGTTGGCAT 2160
QY 2161 GAGGGGAAGAGATGGCTCTCTGTACCAACAAATATTAACAGACTTTTACAGTAAAGAAAT 2220
Db 2161 GAGGGGAAGAGATGGCTCTCTGTACCAACAAATATTAACAGACTTTTACAGTAAAGAAAT 2220

```
QY 2221 GRGGAGGCTCAGATGATTACAAATTTGGTAGGGGCATTTGTATGCCAGTCAACAATT 2280
Db 2221 GTGGCAGGCTCAGATGATTACAAATTTGGTAGGGGCATTTGTATGCCAGTCAACAATT 2280
QY 2281 ATTCTCTTGAAGTAAGGGTTTCGAGCAGGCTTATCATCAATATCCATATCTTGGA 2340
Db 2281 ATTCTCTTGAAGTAAGGGTTTCGAGCAGGCTTATCATCAATATCCATATCTTGGA 2340
QY 2341 GATACATTCATAGGAGTCACAGTTGAAACACATATGAATAATTAATAAAGAGAT 2400
Db 2341 GATACATTCATAGGAGTCACAGTTGAAACACATATGAATAATTAATAAAGAGAT 2400
QY 2401 ATGTTCCCAAGTTCCACAAAGCCAAATACAGATGTGCATTTCTTTTATAAGTCTTCTACA 2460
Db 2401 ATGTTCCCAAGTTCCACAAAGCCAAATACAGATGTGCATTTCTTTTATAAGTCTTCTACA 2460
QY 2461 GCAACACATCTTGATTAATGCGCGATCAACTGCTGTGAAATGAGGTGTAATCTACT 2520
Db 2461 GCAACACATCTTGATTAATGCGCGATCAACTGCTGTGAAATGAGGTGTAATCTACT 2520
QY 2521 AAATCTGGAGCAGGAGTGATTTTCAGTCCCAAGCAAGTGCACAGCAGGTACCTGTGATGG 2580
Db 2521 AAATCTGGAGCAGGAGTGATTTTCAGTCCCAAGCAAGTGCACAGCAGGTACCTGTGATGG 2580
QY 2581 TGTACGTTCTATTTCCTGTGGAGAGTGCTGAAGCTTGCCTCTGTGTACGGAGCATGAC 2640
Db 2581 TGTACGTTCTATTTCCTGTGGAGAGTGCTGAAGCTTGCCTCTGTGTACGGAGCATGAC 2640
QY 2641 TTCCATGAGATTTGAGGAGCCTGCAGAGAGAGATTTTCAGGAAACCTTGTATGTGGAA 2700
Db 2641 TTCCATGAGATTTGAGGAGCCTGCAGAGAGAGATTTTCAGGAAACCTTGTATGTGGAA 2700
QY 2701 GAACCTAAATGCTGATTAAGGAAATTTCTTTGCCCTGAGAAAAGTTGCAACCTCTGAA 2760
Db 2701 GAACCTAAATGCTGATTAAGGAAATTTCTTTGCCCTGAGAAAAGTTGCAACCTCTGAA 2760
QY 2761 ACGGTTGACTTTTGGCTGAAGTGGAGCGGCTGTGGAGCTTTTACTGCCCTTTTGCTG 2820
Db 2761 ACGGTTGACTTTTGGCTGAAGTGGAGCGGCTGTGGAGCTTTTACTGCCCTTTTGCTG 2820
QY 2821 GTGGCTCTGACCTGCTACTTCTTGAAAAGAAATCAAAAGAAAAGTGGTGTAGCCTTGA 2940
Db 2821 GTGGCTCTGACCTGCTACTTCTTGAAAAGAAATCAAAAGAAAAGTGGTGTAGCCTTGA 2940
QY 2881 CTGTTCAACTGAAACCTCAAGATCCCAATATATGAAGACACAGTGTAGCCTTGA 2940
Db 2881 CTGTTCAACTGAAACCTCAAGATCCCAATATATGAAGACACAGTGTAGCCTTGA 2940
QY 2941 GACTAATGAACAAAGAAACCTCTAGTTTACAGGACCATATTTAGGGTCTGTCTTC 3000
Db 2941 GACTAATGAACAAAGAAACCTCTAGTTTACAGGACCATATTTAGGGTCTGTCTTC 3000
QY 3001 ATACCTGTACATTTGGTGATCTACAGAGGAGGCGCATGCCCTGAAAAGGAGAGAT 3060
Db 3001 ATACCTGTACATTTGGTGATCTACAGAGGAGGCGCATGCCCTGAAAAGGAGAGAT 3060
QY 3061 TGAACATTTGATTTGCCTTATCACAATGCTCAAGTACCTTGCCCAATTAAGGAAACCAAT 3120
Db 3061 TGAACATTTGATTTGCCTTATCACAATGCTCAAGTACCTTGCCCAATTAAGGAAACCAAT 3120
QY 3121 GATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAGGAGAGATTTATCTGTATATACAC 3180
Db 3121 GATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAGGAGAGATTTATCTGTATATACAC 3180
QY 3181 ATAACGAAAACCAAGTTTAAGCCCAACCAATGCATGCTGTGATGCAATATAATTAAT 3240
Db 3181 ATAACGAAAACCAAGTTTAAGCCCAACCAATGCATGCTGTGATGCAATATAATTAAT 3240
QY 3241 GGGTAACCTTTTATCTTATGATGTCTACATAACAGTGTGATTTGGAGGACACATGCA 3300
Db 3241 GGGTAACCTTTTATCTTATGATGTCTACATAACAGTGTGATTTGGAGGACACATGCA 3300
```

```
QY 3301 GCATATGCATTTATGATCCAAATTTATCTTTTCTTTGTTTATATATTTTGGGAAAAATAA 3360
Db 3301 GCATATGCATTTATGATCCAAATTTATCTTTTCTTTGTTTATATATTTTGGGAAAAATAA 3360
QY 3361 ATTTTAAAGGTAAATAAAAAA 3390
Db 3361 ATTTTAAAGGTAAATAAAAAA 3390

RESULT 2
US-10-140-164-3
; Sequence 3, Application US/10140164
; Publication No. US20030072736A1
; GENERAL INFORMATION:
; APPLICANT: Baker et al.
; TITLE OF INVENTION: Human Tumor Necrosis Factor Receptor TR16
; FILE REFERENCE: PFS14C1
; CURRENT APPLICATION NUMBER: US/10/140,164
; CURRENT FILING DATE: 2002-05-08
; PRIOR APPLICATION NUMBER: 09/637,856
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/148,348
; PRIOR FILING DATE: 1999-08-12
; PRIOR APPLICATION NUMBER: 60/148,683
; PRIOR FILING DATE: 1999-08-13
; PRIOR APPLICATION NUMBER: 60/148,870
; PRIOR FILING DATE: 1999-08-13
; PRIOR APPLICATION NUMBER: 60/148,758
; PRIOR FILING DATE: 1999-08-16
; PRIOR APPLICATION NUMBER: 60/149,181
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: 60/149,453
; PRIOR FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: 60/149,498
; PRIOR FILING DATE: 1999-08-19
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 3556
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3084)
US-10-140-164-3
```

Query Match 94.8%; Score 3214; DB 9; Length 3556;
Best Local Similarity 95.3%; Pred. No. 0;
Matches 3390; Conservative 0; Mismatches 0; Indels 166; Gaps 1;

```
QY 1 ATGCTGTTCCGCGCCGCGGCGGTACGGGCAGGGGCTGGGGGCGCGCGGGAGGCT 60
Db 1 ATGCTGTTCCGCGCCGCGGCGGCTACGGGCAGGGGCTGGGGGCGCGCGGGAGGCT 60
QY 61 CCCCAGCGCGGCGCTGCGCCGCTCGAGCCCGCCCTGGATTTGCTGCTGGGCGCTCGCC 120
Db 61 CCCCAGCGCGGCGCTGCGCCGCTCGAGCCCGCCCTGGATTTGCTGCTGGGCGCTCGCC 120
QY 121 GCGTCCAGCGGCGCTGGGCTGGGACCTGCGCTCTCTCCAGCGCGCGCTTCCTCT 180
Db 121 GCGTCCAGCGGCGCTGGGCTGGGACCTGCGCTCTCTCCAGCGCGCGCTTCCTCT 180
QY 181 TCCAGGAGAAAGATTTATCATTGTAATATACGGAATGTGATAGCAGTGGCTCCAGGTG 240
Db 181 TCCAGGAGAAAGATTTATCATTGTAATATACGGAATGTGATAGCAGTGGCTCCAGGTG 240
QY 241 AGAGTTGCCATTTCCAAATTTCTGCAGTGGACTGCTCTGGCCTGCCTGACGAGGC 300
Db 241 AGAGTTGCCATTTCCAAATTTCTGCAGTGGACTGCTCTGGCCTGCCTGACGAGGC 300
QY 301 AAAGAATGCATTTCTCTGCTCTTCTGGAGAGTATCTAGAATGAAGAACCGAGTATGC 360
Db 301 AAAGAATGCATTTCTCTGCTCTTCTGGAGAGTATCTAGAATGAAGAACCGAGTATGC 360
```


Db 2521 AAATCTGGAGCAGGAGTGAATTCAGTCCCGCAGCAAGTCCCGCAGGAGTACCTGTGATGGG 2580
QY 2581 TGTACGTTCTATTTCTGTGGGAGAGTGTGAAGCTTCCCTCTGTGTACGAGCATGAC 2640
Db 2581 TGTACGTTCTATTTCTGTGGGAGAGTGTGAAGCTTCCCTCTGTGTACGAGCATGAC 2640
QY 2641 TTCCATGAGATTGAGGAGCCTGCAAGAGAGATTTCAGGAAACCTTGTGTGTGGAA 2700
Db 2641 TTCCATGAGATTGAGGAGCCTGCAAGAGAGATTTCAGGAAACCTTGTGTGTGGAA 2700
QY 2701 GAACCTAATGTGCAATTAAGGAATTTCTTTGCTGTGAGAAAGTTGGCAACCTGTGAA 2760
Db 2701 GAACCTAATGTGCAATTAAGGAATTTCTTTGCTGTGAGAAAGTTGGCAACCTGTGAA 2760
QY 2761 ACGGTTGACTTTTGGCTGAAGTGGAGCGGTGGGAGCTTTTACTGCCGTTTGTG 2820
Db 2761 ACGGTTGACTTTTGGCTGAAGTGGAGCGGTGGGAGCTTTTACTGCCGTTTGTG 2820
QY 2821 GTGGCTCTGACCTGCTACTTCTGAAAAAGAAATCAAAA----- 2858
Db 2821 GTGGCTCTGACCTGCTACTTCTGAAAAAGAAATCAAAAAGTGAATCAAAAATTTCCAG 2880
QY 2859 ----- 2858
Db 2881 TTAGTAATGACGACTAACTCAAAAGAGTGTGAATCCCGCTGCAGACAGTTGTGCTATC 2940
QY 2859 ----- 2858
Db 2941 ATGGAAGGAGAAGATAATGAAGAGGAAGTTGTATATTTCCAATAAACAGTCACTACTAGA 3000
QY 2859 -----GAAAAAGAACCAATTTTGAATCTGTTCACACTGAAA 2894
Db 3001 AAACCTCAAATCTTTGGCAACCAAGGAAAAAGAACCAATTTTGAATCTGTTCACACTGAAA 3060
QY 2895 ACCTCAAGATCCCAATATATGAAGACAGAGTGTGTAGCCCTTGAGACATAATGAACAAA 2954
Db 3061 ACCTCAAGATCCCAATATATGAAGACAGAGTGTGTAGCCCTTGAGACATAATGAACAAA 3120
QY 2955 GAAACCTGCTGTAGTTTACAGGACCATATTTTAGGCTGTCCCTCATACCTGTGCACATT 3014
Db 3121 GAAACCTGCTGTAGTTTACAGGACCATATTTTAGGCTGTCCCTCATACCTGTGCACATT 3180
QY 3015 GGTGATCTCACAGAGAGGGCCATCCCGCTGAAAAGGAGGAGATTGAAACATTTGATT 3074
Db 3181 GGTGATCTCACAGAGAGGGCCATCCCGCTGAAAAGGAGGAGATTGAAACATTTGATT 3240
QY 3075 GCCTTATCACATGGTCAAGTACCTTGCCTGAAAAGGAGGAGATTGTTGGGTCTCAA 3134
Db 3241 GCCTTATCACATGGTCAAGTACCTTGCCTGAAAAGGAGGAGATTGTTGGGTCTCAA 3300
QY 3135 CTGAAGATGAAGCTCAACTCAGGAAGAGATTATCTGTATATACATTAACCTGAAAACCA 3194
Db 3301 CTGAAGATGAAGCTCAACTCAGGAAGAGATTATCTGTATATACATTAACCTGAAAACCA 3360
QY 3195 AGTTTAAAGCCCAACATGACCTGCTGATGCATGCCATATAATTAATGGGTAACCTTTATT 3254
Db 3361 AGTTTAAAGCCCAACATGACCTGCTGATGCATGCCATATAATTAATGGGTAACCTTTATT 3420
QY 3255 CTTTATGATGCTACATAACAAGTGTGATTGGAAGGACATGTGAGCAGATATGCAATATG 3314
Db 3421 CTTTATGATGCTACATAACAAGTGTGATTGGAAGGACATGTGAGCAGATATGCAATATG 3480
QY 3315 ATCCCAATTTATCTTTTCTTTTGTATATTTTGGGAAAAATTAATAATTTTAAAGTA 3374
Db 3481 ATCCCAATTTATCTTTTCTTTTGTATATTTTGGGAAAAATTAATAATTTTAAAGTA 3540
QY 3375 AAAAAAATAAAAAA 3390
Db 3541 AAAAAAATAAAAAA 3556

RESULT 3

us-10-002-050-19
; Sequence 19, Application US/10002050
; Publication No. US20030032095A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corine
; APPLICANT: Yang, Melija
; APPLICANT: Boldog, Ferenc
; APPLICANT: Herrmann, John
; TITLE OF INVENTION: US20030032095A1el Nucleic Acid Sequences Encoding Human Se
; FILE REFERENCE: 15966-554 Cura-54 CON-S14
; CURRENT APPLICATION NUMBER: US/10/002,050
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 09/604,286
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/140,584
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 1737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (296)..(1687)
; NAME/KEY: variation
; LOCATION: (1)..(1737)
; OTHER INFORMATION: N may be any nucleotide
us-10-002-050-19

Query Match 44.2%; Score 1499.6; DB 9; Length 1737;
Best Local Similarity 98.7%; Pred. No. 0;
Matches 1545; Conservative 0; Mismatches 10; Indels 10; Gaps 3;
QY 189 GAAAGATTATCACTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGGAGAGTTGC 248
Db 142 GAAAGATTATCACTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGGAGAGTTGC 201
QY 249 CAITTCAAATTCGAGTGGACTGCTGGCCCTGCTGACCCAGTGAGAGGCAAGAATG 308
Db 202 CAITTCAAATTCGAGTGGACTGCTGGCCCTGCTGACCCAGTGAGAGGCAAGAATG 261
QY 309 CAITTCCTCTGCTTCTGGAGAGTATCTAGAAATGAAGAACCCAGGTATGAGTAAGTG 368
Db 262 CAITTCCTCTGCTTCTGGAGAGTATCTAGAAATGAAGAACCCAGGTATGAGTAAGTG 321
QY 369 TGGTGAAGCACCCTATTCCTGGGAGTGGCATCAAAATTTGATGAATGGATGAATTCGC 428
Db 322 TGGTGAAGCACCCTATTCCTGGGAGTGGCATCAAAATTTGATGAATGGATGAATTCGC 381
QY 429 GCGAGATTTTCTACATCGCAACATTCATGGACACTGTGGTGGGCCCTTCTGACAGCAG 488
Db 382 GCGAGATTTTCTACATCGCAACATTCATGGACACTGTGGTGGGCCCTTCTGACAGCAG 441
QY 489 GCGAGAGGCTGTAACAACCTTCTTGGATCCCTCTGGAACCTACATAGAATCTAATCG 548
Db 442 GCGAGAGGCTGTAACAACCTTCTTGGATCCCTCTGGAACCTACATAGAATCTAATCG 501
QY 549 TGATGACTGCACGGTGTCTTTGATCTATGCTGTGCACCTTAAAGAGTCAGGCTATGCTTT 608
Db 502 TGATGACTGCACGGTGTCTTTGATCTATGCTGTGCACCTTAAAGAGTCAGGCTATGCTTT 561
QY 609 CTTTGAGTACCAGTATGTGCAACAACATCTCTTTTGTAGTCTTTTATTTCAAAATGATCA 668
Db 562 CTTTGAGTACCAGTATGTGCAACAACATCTCTTTTGTAGTCTTTTATTTCAAAATGATCA 621
QY 669 GTGCCAGGAGATGGACACCACCAGTGGGTAAACTTACAGACAATGGAGAATG 728
Db 622 GTGCCAGGAGATGGACACCACCAGTGGGTAAACTTACAGACAATGGAGAATG 681
QY 729 GGGCTCTCATTCTTGTAAATGCTGAAATCAGGCACAAACATACTCTACTGGGAACCTACAGG 788

Db 682 GGGCTCTCAATCTGTAATCTGAAATCAGGCACAAACATACCTACTCTGAGAACTACAGG 741
Qy 789 CATCTCTTATGGTCTTAAGCGGCTCAAGCTGTGCTGGTAAATAATATCACAAATGAAGG 848
Db 742 CATCTCTTATGGTCTTAAGCGGCTCAAGCTGTGCTGGTAAATAATATCACAAATGAAGG 801
Qy 849 GTGGCGGTACACATCAGAAATGTTTCTTGAAGCCAGGCACATTCAGCAACAAACAGG 908
Db 802 GTGGCGGTACACATCAGAAATGTTTCTTGAAGCCAGGCACATTCAGCAACAAACAGG 861
Qy 909 TTCAATCACTGCCAGGTGTCTCCAGAAACACCTATTCTGAGAAAGGAGCCAAAGAATG 968
Db 862 TTCAATCACTGCCAGGTGTCTCCAGAAACACCTATTCTGAGAAAGGAGCCAAAGAATG 921
Qy 969 TATAAGTCTAAGAGGACTCTCAATTTTC-----AGCATCCAGTCACTGTACAGAGG 1022
Db 922 TATAAGTCTAAGAGGACTCTCAATTTTCAGAGGAGGATCCAGTGTGTACAGAGG 981
Qy 1023 CCCTCCCTGTACCAAAAAGACTATTTCAGATCCATCTCCATCTCCATGTGATGAAGAGAA 1082
Db 982 CCCTCCCTGTACCAAAAAGACTATTTCAGATCCATCTCCATGTGATGAAGAGAA 1041
Qy 1083 GACACAGATAATGTACAAGTGTATAGAGCCCAAAATCTCCGGGAGGATCTCACAGATGC 1142
Db 1042 GACACAGATAATGTACAAGTGTATAGAGCCCAAAATCTCCGGGAGGATCTCACAGATGC 1101
Qy 1143 TATTAGATTGCCCTCTTGGAGAGAAGGATTCTCCGCTTGCACCCCTGGATTTTA 1202
Db 1102 TATTAGATTGCCCTCTTGGAGAGAAGGATTCTCCGCTTGCACCCCTGGATTTTA 1161
Qy 1203 TAACAATGGATCATCTCTTGGCCATCCCTGTCTCTCGAAGCATTTTCAGATGGAACCAA 1262
Db 1162 TAACAATGGATCATCTCTTGGCCATCCCTGTCTCTCGAAGCATTTTCAGATGGAACCAA 1221
Qy 1263 AGAATGTAGACCATGTCCAGAGGAGGAGGCTGCACCTTGGCTTTGGAATATAAATGGTG 1322
Db 1222 AGAATGTAGACCATGTCCAGAGGAGGAGGCTGCACCTTGGCTTTGGAATATAAATGGTG 1281
Qy 1323 GAATGTCTCTCTGCGCAACATCAAACTTCCTGCTTCAATGTTGGGAATTCAAAGTGCGA 1382
Db 1282 GAATGTCTCTCTGCGCAACATCAAACTTCCTGCTTCAATGTTGGGAATTCAAAGTGCGA 1341
Qy 1383 TGAATGAATGTTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGGCTTGGAGTTCTGA 1442
Db 1342 TGAATGAATGTTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGGCTTGGAGTTCTGA 1401
Qy 1443 CAATGATTACCTGATCTTAAACTTGCATATCCAGAGTTTAAACCAACCAATCTATGAC 1502
Db 1402 CAATGATTACCTGATCTTAAACTTGCATATCCAGAGTTTAAACCAACCAATCTATGAC 1461
Qy 1503 TGGAGCCAGGGTCTGAACTAGGAAGAATAACATTTGCTTTTGAGACCCCTCTGTTCAGC 1562
Db 1462 TGGAGCCAGGGTCTGAACTAGGAAGAATAACATTTGCTTTTGAGACCCCTCTGTTCAGC 1521
Qy 1563 TGACTGTGTTTGTACTTCAATGGTGGATTAATAGAAAAAGTACAAAATGTGGTAGAATC 1622
Db 1522 TGACTGTGTTTGTACTTCAATGGTGGATTAATAGAAAAAGTACAAAATGTGGTAGAATC 1581
Qy 1623 GTGGGGTGAACCAAGAAAAACAGCTTACACCATATCATCTTCAAGAAATGCAACTTT 1682
Db 1582 GTGGGGTGAACCAAGAAAAACAGCTTACACCATATCATCTTCAAGAAATGCAACTTT 1641
Qy 1683 TACATTTTACATGGG--CATTTCCAGAGACTAAT--CAGGGTCAAGATATACACGGTTCA 1738
Db 1642 TACATTTTACATGGGCAATTCAGAGAACTAATTCAGGGTCCAGATAATAGACGGTTCC 1701
Qy 1739 TCAAT 1743
Db 1702 NCCAT 1706

RESULT 5

US-10-003-152-19
; Sequence 19, Application US/10003152
; Patent No. US20020151494A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corine
; APPLICANT: Yang, Meijia
; APPLICANT: Boldog, Ferenc
; APPLICANT: Herrmann, John
; TITLE OF INVENTION: No. US20020151494A1el Amino Acid Sequences for Human Semaphori
; FILE REFERENCE: 15966-554 Cura-54 CON-S12
; CURRENT APPLICATION NUMBER: US/10/003,152
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 09/604,286
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/140,584
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 1737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (296)..(1687)
; NAME/KEY: Variation
; LOCATION: (1)..(1737)
; OTHER INFORMATION: N may be any nucleotide
US-10-003-152-19

Query Match 44.2%; Score 1499.6; DB 12; Length 1737;
Best Local Similarity 98.7%; Pred. No. 0;
Matches 1545; Conservative 0; Mismatches 10; Indels 10; Gaps 3;

Qy 189 GAAAGATTATCACTTTGAATATACGGAATCTGATACAGTGGCTCCAGGTGGAGAGTTGC 248
Db 142 GAAGATTATCACTTTGAATATACGGAATCTGATACAGTGGCTCCAGGTGGAGAGTTGC 201
Qy 249 CATTCAAATTTGCACTGGAGTGTCTTGGCTTGCCTGACCCAGTGGAGGCAAGAATG 308
Db 202 CATTCAAATTTGCACTGGAGTGTCTTGGCTTGCCTGACCCAGTGGAGGCAAGAATG 261
Qy 309 CACTTTCTCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAACCGATGAGTGAAGT 368
Db 262 CACTTTCTCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAACCGATGAGTGAAGT 321
Qy 369 TGGTGAAGGCACTTATCTCTGGGCACTGGCATCAAAATTTGATGAATGGATGAATGCC 428
Db 322 TGGTGAAGGCACTTATCTCTGGGCACTGGCATCAAAATTTGATGAATGGATGAATGCC 381
Qy 429 GGCAGGATTTTCTAACATCGCAACATTCATGGACACTGTGGTGGGCCCTTCTGACAGCAG 488
Db 382 GGCAGGATTTTCTAACATCGCAACATTCATGGACACTGTGGTGGGCCCTTCTGACAGCAG 441
Qy 489 GCCAGAGCGCTGTAACTCTTCTGGATCCCTCTGGAACACTACATAGAATCTAATCG 548
Db 442 GCCAGAGCGCTGTAACTCTTCTGGATCCCTCTGGAACACTACATAGAATCTAATCG 501
Qy 549 TGATGACTGCAGCGGTGTCTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGCTATGCTTT 608
Db 502 TGATGACTGCAGCGGTGTCTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGCTATGCTTT 561
Qy 609 CTTTGAGTACCAGTATGTGCAACAACATCTTCTTTGAGTTCTTTTATTCAAATGATCA 668
Db 562 CTTTGAGTACCAGTATGTGCAACAACATCTTCTTTGAGTTCTTTTATTCAAATGATCA 621
Qy 669 GTGCCAGGAGATGGACACCACCACTGACAGTGGGTAAAACCTTACAGACAAATGGAGAATG 728
Db 622 GTGCCAGGAGATGGACACCACCACTGACAGTGGGTAAAACCTTACAGACAAATGGAGAATG 681
Qy 729 GGGCTCTCATTCTGTAATGCTGAAATCAGGCACAAACATACCTACTGGGAACACTACAGG 788

Db 682 GGGCTCTCTATTCTGTAATCTGAAATCAGGCACAACTACTCTACTGAGAACTACAGG 741
Qy 789 CATCCCTTATGGTCTTAAGCGGTCAAGCCTGCTGCTGTAATAAATATACAAATTAAGG 848
Db 742 CATCCCTTATGGGTTCTAAGCGGTCAAGCCTGCTGCTGTAATAAATATACAAATTAAGG 801
Qy 849 GGTGCGGTACACATCAGAAATGTTTCCCTTGAAGCCAGGCACATTCAGCAACAAACAGG 908
Db 802 GGTGCGGTACACATCAGAAATGTTTCCCTTGAAGCCAGGCACATTCAGCAACAAACAGG 861
Qy 909 TTCATTCAACTGCCAGGTGTCGCCAGAAACACCTATTCTTGAGAAAGGAGCCAAAGATG 968
Db 862 TTCATTCAACTGCCAGGTGTCGCCAGAAACACCTATTCTTGAGAAAGGAGCCAAAGATG 921
Qy 969 TATAAGGTGTAAGACGACTCTCAATTTTC-----AGCATCCAGTGTACAGAGCG 1022
Db 922 TATAAGGTGTAAGACGACTCTCAATTTTCAGAGGAAGGATCCAGTGTACAGAGCG 981
Qy 1023 CCCTCCCTGTACCAAAAGACTATTTCAGATCCATCTCCATGTGATGAAGAAGGAA 1082
Db 982 CCCTCCCTGTACCAAAAGACTATTTCAGATCCATCTCCATGTGATGAAGAAGGAA 1041
Qy 1083 GACACAGATAATGTACAAGTGAATAGACCCCAAAATCTGCCGGGAGGATCTCACAGATGC 1142
Db 1042 GACACAGATAATGTACAAGTGAATAGACCCCAAAATCTGCCGGGAGGATCTCACAGATGC 1101
Qy 1143 TATTAGATTGCCCTCTCTGAGAGAAGGATGTCGCCCTTCAACCCCTGGATTTTA 1202
Db 1102 TATTAGATTGCCCTCTCTGAGAGAAGGATGTCGCCCTTCAACCCCTGGATTTTA 1161
Qy 1203 TAACAAATGGATCATCTCTTGGCCATCCCTGCTCCTCGAAGCAATTTTCAGATGGAACCA 1262
Db 1162 TAACAAATGGATCATCTCTTGGCCATCCCTGCTCCTCGAAGCAATTTTCAGATGGAACCA 1221
Qy 1263 AGAATCTAGACCATGTCCAGAGGAGGAGGAGGCTGCACCTTGTGATATAAATGGTG 1322
Db 1222 AGAATCTAGACCATGTCCAGAGGAGGAGGAGGCTGCACCTTGTGATATAAATGGTG 1281
Qy 1323 GAATGCTCTTCTGGCAACATGAAACTTCCCTGCTTCAATGTTGGGAATTCAAAGTGCGA 1382
Db 1282 GAATGCTCTTCTGGCAACATGAAACTTCCCTGCTTCAATGTTGGGAATTCAAAGTGCGA 1341
Qy 1383 TGAATGAATGGTGGGAGGTGGCTGGAGATCATATCCAGATGGGGCTGGAGGTTCTGA 1442
Db 1342 TGAATGAATGGTGGGAGGTGGCTGGAGATCATATCCAGATGGGGCTGGAGGTTCTGA 1401
Qy 1443 CAATGATTACCTGATCTTAACTTGCATATCCAGGATTTAAACCCACCAACATCTATGAC 1502
Db 1402 CAATGATTACCTGATCTTAACTTGCATATCCAGGATTTAAACCCACCAACATCTATGAC 1461
Qy 1503 TGGAGCCAGGGTCTGAACCTAGGAAGAATAACATTTGCTTTTGAGACCCCTCTGTTTCAGC 1562
Db 1462 TGGAGCCAGGGTCTGAACCTAGGAAGAATAACATTTGCTTTTGAGACCCCTCTGTTTCAGC 1521
Qy 1563 TGACTGTGTTTGTACTTCAATGATGATTAATAGAAAAGTACAAATGCTGGTAGAATC 1622
Db 1522 TGACTGTGTTTGTACTTCAATGATGATTAATAGAAAAGTACAAATGCTGGTAGAATC 1581
Qy 1623 GTGGGTGGAACCAAGAAAACAAGCTTACACCCATATCATCTTCAAGAAATGCAACTTT 1682
Db 1582 GTGGGTGGAACCAAGAAAACAAGCTTACACCCATATCATCTTCAAGAAATGCAACTTT 1641
Qy 1683 TACATTTACATGGG--CATTTCCAGAGAAGTAAAT--CAGGGTCAAGATAATAGACGGTTCA 1738
Db 1642 TACATTTACATGGGCGATTCCCGAGACAAGTAAATTCAGGGTCCCAAGATAATAGACGGTTCC 1701
Qy 1739 TCAAT 1743
Db 1702 NCCAT 1706
RESULT 6

US-10-002-050-9
; Sequence 9, Application US/10002050
; Publication No. US20030032095A1
; GENERAL INFORMATION:
; APPLICANT: Shinkels, Richard
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corine
; APPLICANT: Yang, Meljia
; APPLICANT: Boldog, Ferenc
; APPLICANT: Herrmann, John
; TITLE OF INVENTION: No. US20030032095A1el Nucleic Acid Sequences Encoding Human Se
; FILE REFERENCE: 15966-554 Cura-54 CON-514
; CURRENT APPLICATION NUMBER: US/10/002,050
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 09/604,286
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/140,584
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1508
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (226)..(1458)
; NAME/KEY: variation
; LOCATION: (1)..(1508)
; OTHER INFORMATION: N may be any nucleotide
US-10-002-050-9

Query Match 34.9%; Score 1182.6; DB 9; Length 1508;
Best Local Similarity 98.4%; Pred. No. 0;

Matches 1228; Conservative 0; Mismatches 10; Indels 10; Gaps 3;

Qy 506 ACTCTTCTGGATCCCTCGTGGAAACTACATAGAACTAATCGTGTGACTGCACGGTCT 565
Db 230 ACTCTTCTGGATCCCTCGTGGAAACTACATAGAACTAATCGTGTGACTGCACGGTCT 289
Qy 566 CTTTGTATCTATGCTGTGCACCTTAAGAACTAGGCTATGCTCTTTTGTAGTACCAATG 625
Db 290 CTTTGTATCTATGCTGTGCACCTTAAGAACTAGGCTATGCTCTTTTGTAGTACCAATG 349
Qy 626 TGCACAACAACATCTTCTTTGTAGTCTTTTATTCAAATGATCAGTCCAGGAGTGCACA 685
Db 350 TGCACAACAACATCTTCTTTGTAGTCTTTTATTCAAATGATCAGTCCAGGAGTGCACA 409
Qy 686 CCACCACTCACAAGTGGGTAAACCTTACAGACAATGGAGAATGGGCTCTCATTTCTGTA 745
Db 410 CCACCACTCACAAGTGGGTAAACCTTACAGACAATGGAGAATGGGCTCTCATTTCTGTA 469
Qy 746 TGCCTGAAATCAGGCACAAACATCTACTTGGAGAACTACAGGCATCTTATGGGTTCTA 805
Db 470 TGCCTGAAATCAGGCACAAACATCTACTTGGAGAACTACAGGCATCTTATGGGTTCTA 529
Qy 806 AGCGGTCAAGCCTGTGCTGGTAAAAATATCACAATTTGAAGGGTGGGCTACACATCAG 865
Db 530 AGCGGTCAAGCCTGTGCTGGTAAAAATATCACAATTTGAAGGGTGGGCTACACATCAG 589
Qy 866 AATGTTTTCTTTCAGAGCCAGGCACATTCAGCAACAAACAGGTTTCATTTCACTGCCAGG 925
Db 590 AATGTTTTCTTTCAGAGCCAGGCACATTCAGCAACAAACAGGTTTCATTTCACTGCCAGG 649
Qy 926 TGTGTCCCAAGAACCTTATCTTGAGAAAGGAGCCAAAGAAATGTAAGGTGTAAAGACG 985
Db 650 TGTGTCCCAAGAACCTTATCTTGAGAAAGGAGCCAAAGAAATGTAAGGTGTAAAGACG 709
Qy 986 ACTCTCAATTTTC-----AGGATCAGTGTACAGAGCCCTTCCCTGTACCAACA 1039
Db 710 ACTCTCAATTTTCAGAGGAGGATCCAGTGTACAGAGCCCTTCCCTGTACCAACA 769
Qy 1040 AAGACTTTTCCAGATCCATCTCCATGTCATGAAGAGGAAAGACACAGATAATGTACA 1099

Db 770 AAGACTATTTCAGATCCATCTCTGATGAAGAAGAAAGACACAGATAATGTACA 829
QY 1100 AGTGGATAGAGCCCAAAATCTGCGGAGGATCTCAGATGCTATTAGATGCCCCCTT 1159
Db 830 AGTGGATAGAGCCCAAAATCTGCGGAGGATCTCAGATGCTATTAGATGCCCCCTT 889
QY 1160 CTGGAGAGAGAGGATGTCTCCGCCCTTGCAACCCCTGGATTTTATAACAATGGATCATCTT 1219
Db 890 CTGGAGAGAGAGGATGTCTCCGCCCTTGCAACCCCTGGATTTTATAACAATGGATCATCTT 949
QY 1220 CTGGCCATCCCTCTCTCTGGAACATTTTCAGATGGAACCAAGAAATGTAGACCATGTC 1279
Db 950 CTGGCCATCCCTCTCTCTGGAACATTTTCAGATGGAACCAAGAAATGTAGACCATGTC 1009
QY 1280 CAGCAGAACGGAGCCTGCACCTTGCGCTTTGAATATAAATGGTGAATGCTCTTCCTGGCA 1339
Db 1010 CAGCAGAACGGAGCCTGCACCTTGCGCTTTGAATATAAATGGTGAATGCTCTTCCTGGCA 1069
QY 1340 ACATGAAACTTCTCTCTCAATGTTGGAAATTCAAAGTCCGATGAATGTTGGG 1399
Db 1070 ACATGAAACTTCTCTCTCAATGTTGGAAATTCAAAGTCCGATGAATGTTGGG 1129
QY 1400 AGTGCTCGAGATCATATCCAGATGGGCTGGAGGTTCTGACAATGATTAACCTGATCT 1459
Db 1130 AGTGCTCGAGATCATATCCAGATGGGCTGGAGGTTCTGACAATGATTAACCTGATCT 1189
QY 1460 TAAACTTGCATATCCAGATTTAAACCAACCAACATCTATGATGAGCCACGGGTTCTG 1519
Db 1190 TAAACTTGCATATCCAGATTTAAACCAACCAACATCTATGATGAGCCACGGGTTCTG 1249
QY 1520 AACTAGAGAAATAACATTTGTTGAGACCTCTGTTGAGCTGACTGTGTTTGTACT 1579
Db 1250 AACTAGAGAAATAACATTTGTTGAGACCTCTGTTGAGCTGACTGTGTTTGTACT 1309
QY 1580 TCATGCTGATATTATAGAAAAGTACAAATGTGGTAGAATCGTGGGTGGAACCAAG 1639
Db 1310 TCATGCTGATATTATAGAAAAGTACAAATGTGGTAGAATCGTGGGTGGAACCAAG 1369
QY 1640 AAAACAAGCTTACAGCCATATCATCTTCAAGAATGCAACTTTTACATTTACATGGG--C 1697
Db 1370 AAAACAAGCTTACAGCCATATCATCTTCAAGAATGCAACTTTTACATTTACATGGGCA 1429
QY 1698 ATTCAGAGAACTAAT--CAGGCTCAAGATAATAGACGGTTTCATCAAT 1743
Db 1430 TTCCCAGAGAACTAATTTACGGGTCCAAGATAATAGACGGTTCCNCCAT 1477

RESULT 7

US-10-002-304-9
; Sequence 9, Application US/10002304
; Publication No. US2003036185A1
; GENERAL INFORMATION:
; APPLICANT: Shinkets, Richard
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corine
; APPLICANT: Yang, Meljia
; APPLICANT: Boldog, Ferenc
; APPLICANT: Herrmann, John
; TITLE OF INVENTION: Polynucleotides and polypeptides encoded thereby
; FILE REFERENCE: 15966-554 Cura-54 CON-58
; CURRENT APPLICATION NUMBER: US/10/002,304
; PRIOR FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 09/604,286
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/140,584
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1508
; TYPE: DNA
; ORGANISM: Homo sapiens

FEATURE:
; NAME/KEY: CDS
; LOCATION: (226)..(1458)
; NAME/KEY: variation
; LOCATION: (1)..(1508)
; OTHER INFORMATION: N may be any nucleotide
US-10-002-304-9

Query Match 34.9%; Score 1182.6; DB 9; Length 1508;
Best Local Similarity 98.4%; Pred. No. 0;
Matches 1228; Conservative 0; Mismatches 10; Indels 10; Gaps 3;

QY 506 ACTCTTCTTTGGATCCCTCGTGGAACTACATAGAATCTAATCTGATGATGACGACGGTGT 565
Db 230 ACTCTTCTTTGGATCCCTCGTGGAACTACATAGAATCTAATCTGATGATGACGACGGTGT 289
QY 566 CTTTGATCTATGCTGTGCACCTTTAAGAAGTCAGGCTATGCTTCTTTTGGTACAGATG 625
Db 290 CTTTGATCTATGCTGTGCACCTTTAAGAAGTCAGGCTATGCTTCTTTTGGTACAGATG 349
QY 626 TCGACAACAACATCTCTTTTGAGTTCTTTATTCAAAATGATCAGTCCAGGAGATGGACA 685
Db 350 TCGACAACAACATCTCTTTTGAGTTCTTTATTCAAAATGATCAGTCCAGGAGATGGACA 409
QY 686 CCACCACTGACAAGTGGTAAACTTTACAGACAATGAGAATGGGGCTCTCATCTCTGTA 745
Db 410 CCACCACTGACAAGTGGTAAACTTTACAGACAATGAGAATGGGGCTCTCATCTCTGTA 469
QY 746 TGCTGAAATCAGGCACAAACATCTCTACTGGAGAACTACAGGCATCTTATGGTCTTA 805
Db 470 TGCTGAAATCAGGCACAAACATCTCTACTGGAGAACTACAGGCATCTTATGGTCTTA 529
QY 806 AGCGGCTCAAGCTGTGCTGTTAAATAATCACAATTTGAAGGGTGGGTACACATCAG 865
Db 530 AGCGGCTCAAGCTGTGCTGTTAAATAATCACAATTTGAAGGGTGGGTACACATCAG 589
QY 866 AATGTTTTCTTGCAGCCAGGCACATTCAGCAACAACAGGTTTCATTCACATGCCAG 925
Db 590 AATGTTTTCTTGCAGCCAGGCACATTCAGCAACAACAGGTTTCATTCACATGCCAG 649
QY 926 TGTGTCCCAGAACACCTATTCTGAGAAAGAGCCAAAGAATGTATAAGGTGTAAAGACG 985
Db 650 TGTGTCCCAGAACACCTATTCTGAGAAAGAGCCAAAGAATGTATAAGGTGTAAAGACG 709
QY 986 ACTCTCAATTTTC-----AGGATCCAGTGAGTACAGAGCCCTCCCTGTACCACAA 1039
Db 710 ACTCTCAATTTTCAGAGGAAGGATCCAGTGAGTACAGAGCCCTCCCTGTACCACAA 769
QY 1040 AAGACTATTTCCAGATCCATCTCTGATGAAGAAGAAAGACACAGATAATGTACA 1099
Db 770 AAGACTATTTCCAGATCCATCTCTGATGAAGAAGAAAGACACAGATAATGTACA 829
QY 1100 AGTGGATAGAGCCCAAAATCTGCGGAGGATCTCAGATGCTATTAGATGCCCCCTT 1159
Db 830 AGTGGATAGAGCCCAAAATCTGCGGAGGATCTCAGATGCTATTAGATGCCCCCTT 889
QY 1160 CTGGAGAGAGAGGATGTCTCCGCCCTTGCAACCCCTGGATTTTATAACAATGGATCATCTT 1219
Db 890 CTGGAGAGAGAGGATGTCTCCGCCCTTGCAACCCCTGGATTTTATAACAATGGATCATCTT 949
QY 1220 CTTGCCATCCCTCTCTCTGGAACATTTTCAGATGGAACCAAGAAATGTAGACCATGTC 1279
Db 950 CTTGCCATCCCTCTCTCTGGAACATTTTCAGATGGAACCAAGAAATGTAGACCATGTC 1009
QY 1280 CAGCAGAACGGAGCCTGCACCTTGCGCTTTGAATATAAATGGTGAATGCTCTTCCTGGCA 1339
Db 1010 CAGCAGAACGGAGCCTGCACCTTGCGCTTTGAATATAAATGGTGAATGCTCTTCCTGGCA 1069
QY 1340 ACATGAAACTTCTCTCTCAATGTTGGAAATTCAAAGTCCGATGAATGTTGGG 1399
Db 1070 ACATGAAACTTCTCTCTCAATGTTGGAAATTCAAAGTCCGATGAATGTTGGG 1129
QY 1400 AGGTGGCTGGAGATCATATCCAGATGGGCTGGAGGTTCTGACAATGATTAACCTGATCT 1459

Db 1130 AGGTGGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTGACAATGATTACCTGATCT 1189
Qy 1460 TAACTTGCATATCCAGAGTTTAAACCACCAACATCTATGACATGGAGCCACGGTTCCTG 1519
Db 1190 TAACTTGCATATCCAGAGTTTAAACCACCAACATCTATGACATGGAGCCACGGTTCCTG 1249
Qy 1520 AACTAGGAAGAAATACATTTCTTTTGGAGCCCTCTGTTTCAAGCTGACATGTTTGTACT 1579
Db 1250 AACTAGGAAGAAATACATTTCTTTTGGAGCCCTCTGTTTCAAGCTGACATGTTTGTACT 1309
Qy 1580 TCATGGTGGATTAATAGAAAAAGTACAAATGCTGGTGAATCTGTTGGGTGGAACCAAG 1639
Db 1310 TCATGGTGGATTAATAGAAAAAGTACAAATGCTGGTGAATCTGTTGGGTGGAACCAAG 1369
Qy 1640 AAAAAACAGCTTACACCCATCATCTTCAAGAATGCAACTTTTACATTTACATGGG--C 1697
Db 1370 AAAAAACAGCTTACACCCATCATCTTCAAGAATGCAACTTTTACATTTACATGGGGCA 1429
Qy 1698 ATTCCAGAGAACTAAT--CAGGGTCAAGATATAGACGGTTCATCAAT 1743
Db 1430 TTCCAGAGAACTAATTCAGGGTCCAAGATATAGACGGTTCNCNCCAT 1477

RESULT 8
US-10-003-152-9
; Sequence 9, Application US/10003152
; Patent No. US20020151494A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corine
; APPLICANT: Yang, Meljia
; APPLICANT: Boldog, Ferenc
; APPLICANT: Hermann, John
; TITLE OF INVENTION: No. US20020151494A1el Amino Acid Sequences for Human Semaphorin-1
; FILE REFERENCE: 15966-554 Cura-54 CON-S12
; CURRENT APPLICATION NUMBER: US/10/003,152
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 09/604,286
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/140,584
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1508
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (226)..(1458)
; NAME/KEY: variation
; LOCATION: (1)..(1508)
; OTHER INFORMATION: N may be any nucleotide
US-10-003-152-9

Query Match 34.9%; Score 1182.6; DB 12; Length 1508;
Best Local Similarity 98.4%; Pred No 0;
Matches 1228; Conservative 0; Mismatches 10; Indels 10; Gaps 3;

Qy 506 ACTCTCTTGGATCCCTCGTGGAACTACATAGAAATCTAATCGTGATGACTGCACGGTGT 565
Db 230 ACTCTCTTGGATCCCTCGTGGAACTACATAGAAATCTAATCGTGATGACTGCACGGTGT 289
Qy 566 CTTTGATCTATGCTGTGCACCTTAAGAACTCAGGCTATGCTCTTTGAGTACCAATG 625
Db 290 CTTTGATCTATGCTGTGCACCTTAAGAACTCAGGCTATGCTCTTTGAGTACCAATG 349
Qy 626 TCGACACCAACATCTCTTTGAGTCTTTTATCAAAATGATCAGTGCCAGGAGATGGACA 685
Db 350 TCGACACCAACATCTCTTTGAGTCTTTTATCAAAATGATCAGTGCCAGGAGATGGACA 409

Qy 586 CCACCACTGACAAGTGGGTAAAACTTACAGACAATGGAGAATGGGCTCTCATTTCTGTAA 745
Db 410 CCACCACTGACAAGTGGGTAAAACTTACAGACAATGGAGAATGGGCTCTCATTTCTGTAA 469
Qy 746 TGCTGAATCAGGCACAAACATCTCTACTGGAGAACTTACAGGCATCTTATGGGTTCTTA 805
Db 470 TGCTGAATCAGGCACAAACATCTCTACTGGAGAACTTACAGGCATCTTATGGGTTCTTA 529
Qy 806 AGGGGGTCAAGCCCTGTGCTGTTAAAAATATACAAATTTGAAGGGGTGGCGTACACATCAG 865
Db 530 AGGGGGTCAAGCCCTGTGCTGTTAAAAATATACAAATTTGAAGGGGTGGCGTACACATCAG 589
Qy 866 AATGTTTTCTTTCAGAGCCAGGCACATTTACAGACAACACAGGTTTCATTTCAACTGCCAGG 925
Db 590 AATGTTTTCTTTCAGAGCCAGGCACATTTACAGACAACACAGGTTTCATTTCAACTGCCAGG 649
Qy 926 TGCTGCCAGAAACACCTTATTTCTGAGAAAGGAGCCAAAGAAATGTATAAGGTGTAAAGACG 985
Db 650 TGCTGCCAGAAACACCTTATTTCTGAGAAAGGAGCCAAAGAAATGTATAAGGTGTAAAGACG 709
Qy 986 ACTCTCAATTTTTC-----AGGATCCAGTGTGTACAGAGGGCCCTCCCTGTACACAA 1039
Db 710 ACTCTCAATTTTTCAGAGGAAGGATCCAGTGTGTACAGAGGGCCCTCCCTGTACACAA 769
Qy 1040 AAGACTATTTCCAGATCCATCTACTCCATGTGTAGAAAGGAAGACACAGATAATGTACA 1099
Db 770 AAGACTATTTCCAGATCCATCTACTCCATGTGTAGAAAGGAAGACACAGATAATGTACA 829
Qy 1100 AGTGGATAGAGCCCAAAATCTGCCGGGAGGATCTCACAGATGCTATAGATTTGCCCCCTT 1159
Db 830 AGTGGATAGAGCCCAAAATCTGCCGGGAGGATCTCACAGATGCTATAGATTTGCCCCCTT 889
Qy 1160 CTGGAGAGAAGAGGATTTGCCCTTGCACCCCTGGATTTTATACAAATGGATCATCTT 1219
Db 890 CTGGAGAGAAGAGGATTTGCCCTTGCACCCCTGGATTTTATACAAATGGATCATCTT 949
Qy 1220 CTTGGCATCCCTGTCTCTCTGCAACATTTTCAGATGGAAACCAAGAAATGTACACCATGTC 1279
Db 950 CTTGGCATCCCTGTCTCTCTGCAACATTTTCAGATGGAAACCAAGAAATGTACACCATGTC 1009
Qy 1280 CAGCAGGAACGGAGCTGCACCTTGGCTTTGAATATAAATGGTGAATGCTTCCCTGGCA 1339
Db 1010 CAGCAGGAACGGAGCTGCACCTTGGCTTTGAATATAAATGGTGAATGCTTCCCTGGCA 1069
Qy 1340 ACATGAAAACTTCTCTGCTTCAATGTTGGGAATTTCAAAAGTGGATGGAATGGAATGGTGG 1399
Db 1070 ACATGAAAACTTCTCTGCTTCAATGTTGGGAATTTCAAAAGTGGATGGAATGGAATGGTGG 1129
Qy 1400 AGGTGGCTGGAGATCATATCCAGATGGGCTGGAGGTTCTGACAATGATTTACCTGATCT 1459
Db 1130 AGGTGGCTGGAGATCATATCCAGATGGGCTGGAGGTTCTGACAATGATTTACCTGATCT 1189
Qy 1460 TAAACTTGCATATCCAGGATTTTAAACCACCAACATCTATGACTGGAGCCACGGTTCCTG 1519
Db 1190 TAAACTTGCATATCCAGGATTTTAAACCACCAACATCTATGACTGGAGCCACGGTTCCTG 1249
Qy 1520 AACTAGGAAGAAATACATTTTGTCTTTTGAGACCCCTCTGTTTCAAGCTGACTGTGTTTGTACT 1579
Db 1250 AACTAGGAAGAAATACATTTTGTCTTTTGAGACCCCTCTGTTTCAAGCTGACTGTGTTTGTACT 1309
Qy 1580 TCATGGTGGATTAATAGAAAAAGTACAAATGCTGTAGAAATGCTGGGTGGAACCAAG 1639
Db 1310 TCATGGTGGATTAATAGAAAAAGTACAAATGCTGTAGAAATGCTGGGTGGAACCAAG 1369
Qy 1640 AAAAAACAGCTTACACCCATCATCTTCAAGAATGCAACTTTTACATTTACATGGG--C 1697
Db 1370 AAAAAACAGCTTACACCCATCATCTTCAAGAATGCAACTTTTACATTTACATGGGGCA 1429
Qy 1698 ATTTCCAGAGAACTAAT--CAGGGTCAAGATATAGACGGTTCATCAAT 1743
Db 1430 TTCCAGAGAACTAATTCAGGGTCCAAGATATAGACGGTTCNCNCCAT 1477

RESULT 9
US-10-028-072-37
; Sequence 37, Application US/10028072
; Publication No. US20030004311A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Collin K
; APPLICANT: Wood, William
; APPLICANT: Zhang
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/028,072
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059836
; PRIOR FILING DATE: 1997-09-24
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/062285
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/062287
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/062814
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/062816
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063045
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063082
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/063127
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063327
; PRIOR FILING DATE: 1997-10-27
; PRIOR APPLICATION NUMBER: 60/063329
; PRIOR FILING DATE: 1997-10-27
; PRIOR APPLICATION NUMBER: 60/063550
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063561
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063704
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/063733
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/063735
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/063738
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/063755
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/064248
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 60/064809
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065846
; PRIOR FILING DATE: 1997-11-17
; PRIOR APPLICATION NUMBER: 60/066364
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/066453
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/066511
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/069212
; PRIOR FILING DATE: 1997-12-11
; PRIOR APPLICATION NUMBER: 60/069278
; PRIOR FILING DATE: 1997-12-11
; PRIOR APPLICATION NUMBER: 60/069334
; PRIOR FILING DATE: 1997-12-11
; PRIOR APPLICATION NUMBER: 60/069694
; PRIOR FILING DATE: 1997-12-16
; PRIOR APPLICATION NUMBER: 60/072320
; PRIOR FILING DATE: 1998-01-23
; PRIOR APPLICATION NUMBER: 60/073612
; PRIOR FILING DATE: 1998-02-04
; PRIOR APPLICATION NUMBER: 60/074086
; PRIOR FILING DATE: 1998-02-09
; PRIOR APPLICATION NUMBER: 60/074092
; PRIOR FILING DATE: 1998-02-09
; PRIOR APPLICATION NUMBER: 60/077791
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079294
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: 60/079663
; PRIOR FILING DATE: 1998-02-27
; PRIOR APPLICATION NUMBER: 60/079728
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/080165
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/081203
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081229
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081695
; PRIOR FILING DATE: 1998-04-14
; PRIOR APPLICATION NUMBER: 60/081817
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081818
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/082999
; PRIOR FILING DATE: 1998-04-24
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/083545
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084627
; PRIOR FILING DATE: 1998-05-07

Qy	231	CTCCAGGTGGAGAGTTGCCATTTCCAAATTCCTGCACTGSACTGCTCTGGCTGCCTGACCC	290
Db	327	TTCCAGGTGGAGGTCGCGCATACCCCGGCGCTGTGCACCAAGCCTGTCTGACCC	386
Qy	291	AGTGAAGGCAAGAANTGCATTTCTCTGTGCTCTTTGGAGAGATCTAGAAATGAAGAA	350
Db	387	CGTGAAGGCCGAGTGTCTCTCTCTGCAAGCGCGGAGGTTCTGGATATGAAGA	446
Qy	351	CCAGGTATGCAGTAAAGTGTGGTGAAGGCAACCTATTCTCTGGCAGTGGCATCAAAATTGGA	410
Db	447	CCAGTCATGTAGCCATTGGCTTGAGGGCGCTACTCCTCGGCACAGGCATTTCGTTTGA	506
Qy	411	TGAATGGGATGAATTTGCCGGCAGGATTTTCTTAAGCATCGCAACATTCATGAGCACTGTGT	470
Db	507	TGAGTGGGATGAGCTGCCCATATGGCTTTTGGCAGGCTCTCAAGCCAAATGAGAGCTGGATGA	566
Qy	471	GGGCCCTTCTGACAGCAGGCGAGCGGTGTAAACAATCTTCTTTGGATCCCTCGTGGAAA	530
Db	567	CAGTGTCTGAGTGCA---CGGGAACTGTACTTGTCTCCAAGTGGGTTCCCGGGCGCA	623
Qy	531	CTACATAGAATCTAATCGTGATGACTGCACGGTGTCTTTGATCTATGCTGTGCACCTTAA	590
Db	624	CTACATCGCTCCAAACACGGAGAAATGCACAGCCACACTGATGTACCGCTCAACCTGAA	683
Qy	591	GAAGTCAGGCTATGCTCTCTTGTGAGTACCAGATGTGCGACAACAACATCTCTTTGAGTT	650
Db	684	GCAATCTGGCACCCTTAACTTCGAACTACTATCCAGCTCCAGCATCATCTTTGAGTT	743
Qy	651	CTTTATTTCAAATATGATCGCTGCCAGGAGATGSACACCACTGACAAGTGGTAAAACT	710
Db	744	TTTCTGTTCAGAAATGACAGTGCAGCCCAATGACATGACTC---CAGGTGGATGAAGAC	800
Qy	711	TACAGACAATGGAGATGGGGCTCTCATCTGTATGCTGAATCAGGCACAAACATACT	770
Db	801	CACAGAGAA---AGSATGGGAATCCACAGTGTGGAGTAAATCGAGCGCAATATGTCT	857
Qy	771	CTACTGCAGAACTACAGGCATCTTATGGGTTCTAAGCGGTCAAGGCTGTGTGTGTAAA	830
Db	858	CTATTGGAGAACCACAGCCTTCTCAGTATGGACCAAAAGTACCCAAGCCTGTGTGTGTAG	917
Qy	831	AAATATCACAAATTTGAAGGGTGGCTTACACATCAGAAATGTTTTCTTTGCAAGCCAGGCAC	890
Db	918	AAACATTTGCCATAACAGGGTGGCTTACACTTCAGAATGCTTCCCTCTGCAAACTGCGCAC	977
Qy	891	ATTTCAGCAACAACACAGGTTTCACTTAAGTGTGTCCTCCAGAAACACCTATTCTGA	950
Db	978	GTATGACAGACAGAGGGTCTCTTTCTGCAAACTTTGCCCGCCAACTCTATTTCAAA	1037
Qy	951	GAAGGAGCCAAAGAATGTATAGGTGTAAA---GACGACTCTCAATTTTCAGATCCAG	1007
Db	1038	TAAAGGAGAACTTCTTGCCACCGTGTGACCTTGACAAATACTCAGAGAAAGATCTTC	1097
Qy	1008	TGAGTGTACAGAGGCCCTCCCTGTACCACAAAAGACTATTTCAGATCCATCTCCATG	1067
Db	1098	TTCTCTGTAAAGCTGGGCCAGTTTGCACAGACAAGATATTTCATACACACAGCGGCTG	1157
Qy	1068	TGATGAAGAAGAAAGACAGATAATGTACAAGTGGATAGAGGCCCAAAATCTGCCGGGA	1127
Db	1158	CGATGCCAACGGAGACACAACCTCATGTACAAATGGCCAAAGCCGAAATCTGTAGCGA	1217
Qy	1128	GGATCTCACAGATGCTATTAGATTGCCCTTCTGGAGAGAAGAGGATGTCTCGCCCTTG	1187
Db	1218	GGACTTTGAGGGGCACTGAAGCTGCTGCTCTCTGGTGTGAAGACCCACTGCCCAACCTG	1277
Qy	1188	CAACCTCGGATTTTATAACAATGGATCATCTTCTTTGCCATCCCTGTCTCTCTGGAACATT	1247
Db	1278	CAACCCAGGCTTCTTCAAAACCAACAACAGCACCTGCCAGCCTGCCCATATGTTCTCTA	1337
Qy	1248	TTTCAGATGGAAACCAAGAATGTAGACCATGTCTCAGCAGACGGAGCCTGCACATTGGCTT	1307
Db	1338	CTCCAAATGG---CTTCAGACTGTACCCGCTGCCTCGCCTGCAGGGAATGAACCTGCTGTGGGAT	1394

Qy	1308	TGAATATAAATGGTGAATATGCTCTCTGGGCACATGAAACACTTCCCTGCTTCAATGTTGG	1363
Db	1395	TGAATACAAATGGTGGAAACAGCCTGCCACAAACATGAAAGACACCGTCTCAGTTGGGAT	1454
Qy	1368	GAATTCAAAGTGGGATGGAATGAMTGGTTGGGAGGTGGCTGGAGATCATATCCAGAGTGG	1427
Db	1455	CACTTTCAGTACAAAGGCATGACAGGTGGGAGGTGGCTGGTGTATCACAATTTACACAGC	1514
Qy	1428	GGCTGGAGGTTCTGACAAATGATTAACCTGATCTTAAACTTTGGATATCCCAAGATTTAAACC	1487
Db	1515	TGCTGGAGCCTCAGACAAATGACTTTCATGATTTCTACTCTGTTGTGGCCAGGATTTAGACC	1574
Qy	1488	ACCAACATCTATCACTTGGAGCCAC--GGGTCTCAACTAGGAGAAATTAACATTTTGTCTT	1544
Db	1575	TCCGAGCTCGGTGATGGCAGACAGAGANATAAGAGGTGGCCAGATTCACATTTTGTCTT	1634
Qy	1545	TGAGACCTCTGTTCAGCTGACTGTGTTTGTACTTTCATGTGGTATTAATAGAAANAAG	1604
Db	1635	TGAGACCTCTGTCTGTGAATCTGTGAGCTCTACATTTCTATGGTGGTGTGAATTTCTAGGAC	1694
Qy	1605	TACAAATGCTGGTAGAATCGTGGGCTGGAACCAAGAAAGAAACANAGCTTACACCCATATCAT	1664
Db	1695	CAACACTCTCTGGAGCGTGGAAAGGTTCCAAAGGCAACAGTCTTATACCTACATCAT	1754
Qy	1665	CTTCAAGAAATGCAACTTTTACATTTACATGGGCATTTCCAGAGAACTTAATCAGGTCGAAGA	1724
Db	1755	TGAGGAGAACACTTACCACGAGCTTCACTTGGCCCTTCCAGAGGACCACCTTTTTCATGAGGC	1814
Qy	1725	TAATAGAGGTTTCATANTGACATGTTGAAGATTTATTTCTATCACAGCCACTAATGCACT	1784
Db	1815	AAGCAGGAAGTACACCAATGACCTTGCCAAGATCTACTCCATCAATGTACCAATTTTAT	1874
Qy	1785	TGATGGGTTGGCTCTCATGCGCTGCTGCGCTCTGAAACAGTCCGGTTTCATC	1844
Db	1875	GAATGGGCTGGCTCTACTGCGCTCCCTGTGCGCTAGAGCCCTCTGATGTGGGCTCCTC	1934
Qy	1845	GTGTGTCCCCTGGCCCTCCAGGCCACTTACATTTGAGAAAGAAACCAACAGTGCAGAGGATG	1904
Db	1935	CTGCACCTCTTGTCTGCTGTTACTATATTTGACCGAGATTTCAGGAACCTGCCACTCGTG	1994
Qy	1905	TCCACTTGACACCTACCTGTCCATACATCAGGCTTATGCGAAAGAGGTTGTTATTTCCATG	1964
Db	1995	CCCCCTTAACACAAATTTCTGAAGCCCAACAGCCCTTATGCTGCCAGGCTGTGTGCCCTG	2054
Qy	1965	CGGCGCTGGGAGTAAACAAATCAGGACCATTTGGTTTGTCTATGTGACTGCTTTTTCATA	2024
Db	2055	TGGTCCAGGGACCAAGAAACAAGATCCACTCTCTGTGCTACAATGATTCGACCTTCTC	2114
Qy	2025	CCATTGAANAAGAAATCAGATTTTGGACATATGACTTATAGCAACCTTACGACGTGTGGGCTC	2084
Db	2115	ACGCAACACTTCAACAGGACTTTCAACTACAACTTCTCCGGTTTGGCAAAACACCGTCAAC	2174
Qy	2085	ATTATTAAGTGGCCGCCAGCTTCACTTCCAAAGGAACAAATAACTTTCCATTTCTCAATAT	2144
Db	2175	TC TTGCTGGAGGCCAAGCTTCACTTCCAAAGGTTGAAATACTTTCCATCTATCCCTT	2234
Qy	2145	CAGTTTATGTGGGCATGAGGGGAAGAGATGGCTCTCTGTACCAACATATAACAGACTT	2204
Db	2235	CAGTCTCTGTGGAACACAGGTTAGGAAAATGTCTGTGTCCACCCGACATGTCACTGACCT	2294
Qy	2205	TACAGTAAAGAAATAGTGGCAGGCTCAGATGATTAACAAAAATTTGGTAGGGGCAATTTGT	2264
Db	2295	CCGGATTTCTGAGGGTGAGTCAGG-----TTC TCCAATCTATCACAGCCTACGT	2345
Qy	2265	ATGCCAGTCAACAATATTTCTTCTGAAAGTAAAGGTTTCCGAGCAGCCTTATCATCAACA	2324
Db	2346	CTGCCAGGCACTCATCATCTCCCCACAGAGGTGACAGGCTACAAGGGGGGGTTTCTCTCA	2405
Qy	2325	ATCCATCATTTCTGGCAGATCATTTATAGGAGTACAGTTTGAAGCCACATTTGAAAATAT	2384
Db	2406	GCCTGTGAGCCTTCTGATCGACTTTTGGGGTGACACAGATATGACTCTGGATGGAAAT	2465
Qy	2385	TAATATAAAGAAAGATATGTTCCCAAGTTCCACAAAGCCAAATACCAAGATGTGCATTTCTTT	2444

Db 2466 CACCTCCACAGTGAACCTTTCCACCTGGAGTCCTTGGGAATACCGACGTCATCTCTT 2525
Qy 2445 TTATAAGTCTCTACAGCAACAACATCTGTATTAAATGCGCCGATCAACTGCTGTGAAT 2504
Db 2526 TTATAGTCCCAATGATGACCCAGTCTCTGCAAGTCTGCGGAGATCAACACCATCGCGT 2585
Qy 2505 GAGGTGAATCTACTAAATCTGGAGCAGAGTGAATTTTCAGTCCCGCAGCAAGTGCACAGC 2564
Db 2586 CAGGTGCAGTCCACAGAAAACCTGTCCTGGAAGTTTGCTGCTGCCAGGAACCTGCTCAGA 2645
Qy 2565 AGTACCTGTGATGGGTGTACCTTCTATTCTCTGCGGAGAGTGTGAAGCTTGCCCTCT 2624
Db 2646 TGGGACCTGTGATGGCTGCAACTTCCACTTCTCTGCGGAGAGCGCGCTGCTTGCCCGCT 2705
Qy 2625 GTGTGAGGAGTGAATCTCCATGAGATTCAGGAGCGCTGCAAGAGAGGATTTTCAGGAAC 2684
Db 2706 CTGCTCAGTGGCTGACTACCATGCTATCTGTCAGAGCTGTGTGGTGGGATCCAGANGAC 2765
Qy 2685 CTGTATGTGTGGAATGAACCTAAATGTGTGCAATTAAGGAATTTCTTTCGCTTGAGAAAA 2744
Db 2766 TACTTACGTGTGNCAGAACCCCAAGCTATGCTCTGTGTGSCATTTCTCTGCTGAGCAGAG 2825
Qy 2745 GTTGGCAACCTGTGAACGGTGTACTTTTGGCTGAAGGTGGAGCGGTGTGGGAGCTTT 2804
Db 2826 AGTCACCATCTGCAAAACCATAGATTTCTGGCTGAAAGTGGGCATCTCTGCGAGGACCTG 2885
Qy 2805 TACTGCGGTTTGTGGTGGCTGCTGACCTGCTACTTCTGGAAGAAAGATCAAAAGAAAA 2864
Db 2886 TACTGCCATCTGCTCAGCGTCTGACCTGCTACTTTTGGAAAAAGATCAAAAACTAGA 2945
Qy 2865 GAAGACCAATTTTGAATCTGTCTCA 2887
Db 2946 GTACAAGTACTCCAGCTGGTCA 2968

RESULT 11

US-10-123-904-37
; Sequence 37, Application US/10123904
; Publication No. US20030022328A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330RIC54
; CURRENT APPLICATION NUMBER: US/10/123,904
; PRIORITY FILING DATE: 2002-04-16
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 37
; LENGTH: 3501
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 2762, 2778
; OTHER INFORMATION: unknown base

US-10-123-904-37

Query Match 26.6%; Score 901; DB 9; Length 3501;
Best Local Similarity 60.0%; Pred. No. 2.5e-241;
Matches 1634; Conservative 0; Mismatches 1062; Indels 27; Gaps 7;
Qy 171 GCTTCTCTCTTCCAGGAGAAAAGATTATCATCTTTGAATATAGGAATGTGATAGCAGTG 230
Db 267 GCTTCATGCTGCAAGAGTCTGAGTACCACATATGAGTACACGGCTGTGACAGCAGCGG 326
Qy 231 CTCAGGTGGAGAGTTGCCATTCCAAATTTCTGCAGTGGACTGCTTGGCCTGCCTGACCC 290
Db 327 TTCCAGGTGGAGGTGCGCGTCCGCATACCCCGGCCCTGTGCACAGCCTGTCTGACCC 386
Qy 291 AGTGAGAGCAAGAGATGCACCTTCTCCTGTCTCTTGGAGAGTATCTAGAATAGAA 350
Db 387 CGTCAAGGCAACCGAGTGTCTCTTCTTCCGCAACGCGCGGGAGTTTCTGGATATGAAGA 446
Qy 351 CCAGGTATGCAGTAAGTGTGTGAAGGCACCTATTCTTGGGCGAGTGGCATCAAAATTGA 410
Db 447 CCAGTCATGTAAGCCATGCGCTGAGGCGGCTACTCCCTCGGCACAGGCATTCGGTTTGA 506
Qy 411 TGAATGGGATGAATTCGCGGAGGATTTTCTTAACATCGCAACATTCATGGACACTCTGTGT 470
Db 507 TGAGTGGGATGAGTGCCTCATGGCTTTGCCAGCCTCTCAGCACAACATGGAGCTGATGA 566
Qy 471 GGGCCTTCTGACAGCAGCGCCAGCGCTGTAAACACTCTTCTTGGATCCCTCGTGAA 530
Db 567 CAGTGTCTGTAGTCCA---CCGGGAACCTGTACTTGTCTCAAGTGGGTTCCTCCGGGCGA 623
Qy 531 CTACATAGAAATCTAATCGTGATGACTGACCGGTGTCTTTGATGATCTATGCTGTGCACCTAA 590
Db 624 CTACATCGCTCCACACGCGAGATGACACGCCACTGATGTAGCGCTCAACCTGAA 683
Qy 591 GAAGTCAGCTATGCTTCTTTGAGTACAGTATGTGACAAACATCTTCTTTGAGTT 650
Db 684 GCAATCTGGCACCCTTAACCTTCGAATCTACTTCCAGACTCCAGCATCATCTTTGAGTT 743
Qy 651 CTTTATTCAAATGATCAGTGGCAGGAGATGCACACCACCATGACAAAGTGGTAAACT 710
Db 744 TTTGTTTCAAGTACAGTGGCAGCCCAATGCAGATGACTC---CAGGTGATGAAGAC 800
Qy 711 TACAGACAATGGAGTGGGCTCTCATCTGTGAATCTGAAATCAGGCACAAACATCT 770
Db 801 CACAGAA---AGGTGGGAATTCACAGTGTGGAGCTAAATCGAGGCAATAATCTCT 857
Qy 771 CTACTGGAGAACTACAGGCATCTTTATGGTGTCTAAGCGGTCAAGCCTGTGCTGTAA 830
Db 858 CTATTGGAGAACACAGCCTTCTCAGTATGGACCAAGTACCAAGCCTGTGCTGTAG 917
Qy 831 AAATATCAATTTAAGGGTGGCTACACATCAGAAATGTTTCTTTCGAAGCCAGGCAC 890
Db 918 AAACATTTGCCATAACAGGGTGGCTTACACTTCAGAAATGCTTCCCTGCAAACTCTGCAC 977
Qy 891 ATTCAGCAACAAACAGGTTTCAATTCAGTGGCAGGTTGCTCCAGAAACACCTATTCTGA 950
Db 978 GTATCAGACACAGCAGGCTCTCTTTCTGCAAACTTTGCCAGCCTCACTTTATTCAA 1037
Qy 951 GAAAGAGCAAGAAATGTATAGGTGTA---GACGACTCTCAATTTTCAGGATCCAG 1007
Db 1038 TAAAGGAGAACTTTCTGACCCAGTGTGACCTTGCAAAATCTCAGAGAAAGGATCTTTC 1097
Qy 1008 TGAGTGTACAGAGCCCTCTCTGTACCAACAAAGACTATTTCCAGATCCATCTCATG 1067
Db 1098 TTTCTGTAACTGTGCGCCAGTGTGACAGACAAAGATTTATTTACACACACACGCGCTG 1157
Qy 1068 TGATGAAGAGAAAGACACAGATATGTATCAAGTGGATAGAGCCCAAAATCTGCGGGA 1127
Db 1158 CGATGCCACGAGAGACACACTCATGTACAAATGGGCGCAAGAAATCTGTAGCA 1217
Qy 1128 GGATCTCAGAGTGTATTAGATTGCCCTTCTTGAGAGAAAGAGATTGTCGCGCTTG 1187
Db 1218 GGACCTTGAGGGGAGTCAAGCTGCCTCTGCTGTGTGAAGACCCACTGCCACCCCTG 1277

Qy	1188	CAACCCCTGGATTTTATAACAANTGGATCATCTTCTTTGCCATCCCTGTCTCTCGAACAATT	1247
Db	1278	C AACCCAGGCTTCTTCAAAACCACAACAGCACCTGCACGCCCTGCCCATATGTTTCCCTA	1337
Qy	1248	TTCAGATGGAAACAAAGAATGTAGACCATGTCCCAGCAGGAACGGAGCGCTGCACATTGGCTTT	1307
Db	1338	CTCCAATGG---CTCAGACTGTACCCCGTGCCCTGCAGGGACTGAACCTGCTGTGGGAA TT	1394
Qy	1308	TGAATATAAATGGTGGAAATGCTCTTCTTGGCAACATGAAAACTTCCCTGCTTTCAATGTTGG	1367
Db	1395	TGAATACAAATGGTGGAACACGCTGCCACAACATGTGAAACGACCGTCTCTCAGTGGGAT	1454
Qy	1368	GAATTCAAAGTCGGATGGAAATGAAATGGTTGGAGGTGGCTGGAGATCATATPCAGAGATGG	1427
Db	1455	C AACTTCAGATACAAGGGATGACAGGCTGGGAGGTGGCTGTGATCACATTTACACAGC	1514
Qy	1428	GGCTGGAGGTTCTGACAAATGATTACCTGATCTTAACCTTGGCATATCCCAGGATTTAAACC	1487
Db	1515	TGCTGGAGCCTCAGACAAATGACTTCATGATTTCTCACTCTGTTGTGCCAGGATTTAGACC	1574
Qy	1488	ACCAACATCTATGACTGGAGCCAC---GGGTCTTCAACTAGGAAGAATAACATTTGTCTTT	1544
Db	1575	TCCGACGTGCGTATGSCAGACACAGAGAATAAGAGGTGGCCAGAATCACATTTGTCTTT	1634
Qy	1545	TGAGACCCCTGTTTACGCTACGCTGTGTTTTGTGTACTTTTCATGTGTGGATATTAATAGAAAAAG	1604
Db	1635	TGAGACCCCTCTGTTCTGTGAACGTGTGAGCTCTACTTTCATGTGTGGTGTGAAATCTTAGGAC	1694
Qy	1605	TACAAATGTGGTAGAATCGTGGGCTGGAACCAACAAACAAAACAAAGCTTACACCCATATCAT	1664
Db	1695	CAACACTCCTGTGGAGCCTGGAAAGTTCCAAAGGCCAAACAGTCTTACCTACATCAT	1754
Qy	1665	CTTCAAGAATGCAACTTTTACATTTTACATPGGCATTCACAGAGAACTAATCAGGCTCAAGA	1724
Db	1755	TGAGGAGAACACTTACCACGAGCTTCACTTGGGCCCTTCCAGAGGACCACTTTTCATGAGGC	1814
Qy	1725	TAATAGACGTTTCATCAATACATGTTGAAGATTTATTTCTATCACGCCACTANTGCAGT	1784
Db	1815	AAGCAGGAAGTACACCAATGAGCTTGCCAAAGATCTACTTCCATCAATGTACCAATGTTAT	1874
Qy	1785	TGATGGGTGGCCTCTCATGGCGTGCCTGTGCCCTCTCGGTTCTGAAACAGTCGGGTTCATC	1844
Db	1875	GAATGGCGTGGCCTCTACTTGGCGTCCCTGTGCCCTAGAACCTCTGTATGTGGGCTCCTC	1934
Qy	1845	GTGTGTCCTTGGCCTCCAGGCCACTACATTTGAGAAAGAAACCAACAGTGCAGGAAGATG	1904
Db	1935	CTGCACCTCTGTCTGCTGGTTACTATATTTTACGCGAGATTCAGGAACCTGCCACTCC TG	1994
Qy	1905	TCCACCTTGACACCTTACTGTCCATACATCAGTCTATGSAAGAGGCTGTATTTCCTCATG	1964
Db	1995	CCCCCTTAACACAATTTCTGAAAGCCCAACGCTTATGGTGTCCAGGCTGTGTGGCCCTG	2054
Qy	1965	CGGCGCTGGGAGTAAAACAANTCAGGACCATTTGGTTTGCCTATAGTACTGCTTTTTCCTA	2024
Db	2055	TGGTCCAGGGACCAAGAACAAGAATCCACTCTGTGCTCAATGATGTCACCTTCTC	2114
Qy	2025	CCATGAAAAAGAAATACGATTTTGGCATATGACTTTTAGCAACCTCAGCAGTGTGGGCTC	2084
Db	2115	ACGCAACACTTCCAACGAGCACTTCAACTTACAACTTCTCCGCTTTGGCAACACGCTGCAC	2174
Qy	2085	ATTAAATGAATGGCCCGCAGCTTCACTTCCAAGGAACAATACTTCCAATTTCTCAATAT	2144
Db	2175	TC TGTCTGGAGGCCAANGCTTCACTTCCANAGGTTGAAATACATTCACTTACCTTTAC CCT	2234
Qy	2145	CAGTTTATGTGGGCATGAGGGGAAGAAGATGCTCTCTGTACCAACAATATAACAGACTT	2204
Db	2235	CAGTCTCTGTGGAAACAGSGTAGGAAATGCTCTGTGTGCACCCGACAAATGTCTACTGACCT	2294
Qy	2205	TACAGTAAAAGAAATAGTGGCAGGGTFCAGATGATTTACACAAATTTGGTAGGGCAATTTGT	2264
Db	2295	CCGGATTTCTGTAGGGTGTGCTCAGG-----TCTTCCAAATCTATCACAGCCTACGT	2345

Qy	23265	ATGCCAGTCAACAATTTATTTCTCTGTAAAGTTAAGGGTTTCCGAGGAGCCTTATCATCACA	23262
Db	23436	CTGCCAGGCAGTCATCATCCCCCAGAGGTGACAGGGCTACAAGCGGGGGTTTCCCTCACA	24005
Qy	23225	ATCCATCATTTCTGGCAGATACATTTCATAGGAGTACAGTTGAAACACCATTTGAAANAATAT	23884
Db	24006	GCCTTGTCAGCCTTGCCTGATCGACTTATTGGGGTGTGACAACAGATATGACTCTGGATGCAAT	24665
Qy	23855	TAATATAAAGAAGATATGTTCCCGAGTTCCACACAGCCAAATACCNAGATGTGCATTTCTTT	24444
Db	24666	CACCTCCCGAGCTGAACCTTTTCCACCTCGAGGTCTTTGGGAATACCGGACGTGATCTTCTTT	25225
Qy	24445	TTATAAGTCTTTCTACAGCAACAACATCTTTGTATTAAATGGCCGATCAACTGCTGTGAAAAAT	25004
Db	25226	TTATAGGTCCAATGATGTACCCAGTCTCTGCGAGTTCTGGGAGATCAACCCACATCTCCGCT	25885
Qy	25005	GAGGTGTAATCTTACTTAATCTTGGAGCAGGAGTGTATTTCACTGCCCAAGCAAGTCCCGCAGC	25664
Db	25886	CAGGTGCAGTCCACAGAAACATGTCCCTGGAGTTTGTCTGCTGCCAGGACGTCTCCAGA	26445
Qy	25665	AGGTACTGTGATGGGTGTACGTTCTATTCTCTGGGAGAGTGTCTGAAGCTTGCCCTCTCT	26224
Db	26446	TGGGACCTGTGTATGGGTGCAACTTTCACATCTCTGTGGGAGAGCGGGTGTCTGCCCGCT	27005
Qy	26225	GTGTACGGAGCATGACTTCCATCAGATTTGAGGGAGCCTTGCAAGAGAGGATTTTCAGGAAC	26884
Db	27006	CTGCTCAGTGGCTACTACCATGCTATCGTACAGAGCTGTGTGGCTGGGATCCAGANGAC	27665
Qy	26885	CTTGTATGTGTGGAAATGAACCTTAATGTGTGCATTTAAAGGAATTTCTTTTGCTTGAGAAAA	27444
Db	27666	TACTTACTGTGTGNCAGAGAACCCAAAGCTATGCTCTGGTGGCATTTCTCTGCTCTGAGCAGAG	28225
Qy	27445	GTTTGGCAACCTGTGAAACCGTTTGACTTTTGGCTGAAGGTGGGAGCGGCTGTGGGAGCTTTT	28004
Db	28226	AGTCACCATCTGCANAACCATAGATTTCTGGCTGAAAGTGGGCATCTCTGCGAGCACCTG	28885
Qy	28005	TACTGGCGTTTCTCTGGTGGCTCTGACCTGCTTACTTCTTGGAAAGAAGATCAAAAGAAAA	28684
Db	28886	TACTGGCATCCTCTCTCACCGTCTTTGACCTGCTTACTTTTGGAAAGAAGATCAAAACATAGA	29445
Qy	28665	GAAGACCATTTTGAATCTGTCTCA	2887
Db	29446	GTACAAGTACTCCAAAGCTGTGA	2968

RESULT 12

RESOLI 12
US-10-140-470-37

US-10-140-470-37
; Sequence 37, Application US/10140470

; sequence 37, Application 03/101
; Publication No. US20030022331A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.

APPLICANT: Beresini, Maureen

APPLICANT: DeForge, Laura

APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ell

APPLICANT: Gao, Wei-Qiang

APPLICANT: Gerritsen, Mary E

APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy

APPLICANT: Tumas, Daniel

APPLICANT: Watanabe, Colin K
APPLICANT: Wood William

APPLICANT: Wood, William

APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRET

SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME

; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C160; FILE REFERENCE: P3330RICI60
: CURRENT APPLICATION NUMBER: 0

```

; CURRENT APPLICATION NUMBER: US/10/140,470
; CURRENT FILING DATE: 2002-05-06

```

; CURRENT FILING DATE: 2002-03-06
 ; Prior Application removed - See

, PILOT APPLICATION REMOVED - SEE PART III OF FIRE WRAPPER

APPLICANT: Gao,Wei-Qiang
APPLICANT: Gerritsen,Mary E.
APPLICANT: Goddard,Audrey
APPLICANT: Godowski,Paul J.
APPLICANT: Gurney,Austin L.
APPLICANT: Sherwood,Steven
APPLICANT: Smith,Victoria
APPLICANT: Stewart,Timothy A.
APPLICANT: Tumas,Daniel
APPLICANT: Watanabe,Colin K
APPLICANT: Wood,William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P33301C382
CURRENT APPLICATION NUMBER: US/10/176,918
CURRENT FILING DATE: 2002-06-20
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 37
LENGTH: 3501
TYPE: DNA
ORGANISM: Homo Sapien
FEATURE:
NAME/KEY: unsure
LOCATION: 2762, 2778
OTHER INFORMATION: unknown base
US-10-176-918-37

Query Match 26.6%; Score 901; DB 9; Length 3501;
Best Local Similarity 60.0%; Pred. No. 2.5e-241;
Matches 1634; Conservative 0; Mismatches 1062; Indels 27; Gaps 7;
Qy 171 GCTTCTCTCTCCAGGAGAAAGATTATACATTTGATATACGGAATGTAGTAGAGTGG 230
Db 267 GCTTCAGCTGCAAGAGCTGAGTACCACATATGATACAGCGGTGTGACAGCAGGG 326
Qy 231 CTCAGGTGGAGAGTTGCCATTTCCAAATTTCTGCAGTGGACTGCTCTGGCTCCCTGACCC 290
Db 327 TTCCAGGTGGAGGGTGGCGTGGCCCATACCCCGGCGCTGTGCACAGCCCTCTCTGACCC 386
Qy 291 AGTGAGAGCAAGAAGTACGATTTCTCTCTGCTCTCTGAGAGATCTAGAAATGAAGAA 350
Db 387 GGTCAAGGCCACCGAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 446
Qy 351 CCAGGTATGCAATGAGTGTGGTGAAGGCAACCTATTCTTTGGGAGTGGCATCAAAATTGA 410
Db 447 CCAGTCATGTAAGCCATGGCTGAGGGCGCTACTCCCTCGGCACAGGSCATTCGGTTGA 506
Qy 411 TCAATGGGATGAATTTCCGGGCAAGGATTTTCTAACATCGCAACATTCATGGACACATGTGT 470
Db 507 TGAGTGGGATGAGTGTCCCATGGCTTTGCCAGCCCTCTCAGCCAACTGGAGCTGGATGA 566
Qy 471 GGGCCCTCTGACAGCAGGCCAGAGCGGTGTAAACAATCTCTTGGATCCCTCGTGGAA 530
Db 567 CAGTGTCTGAGTCCA---CCGGNACTGTACTTCGTCCAAGTGGTTCCTCCCGGGGCGA 623
Qy 531 CTACATAGAATCTAATCTGTGACTGACGAGTGTCTTTGATCTATGCTGTGACACCTTAA 590
Db 624 CTACATCGCTCCAAACAGCGAGCAATGACACGCCACACTGATGACGCGTCAACCTGAA 683
Qy 591 GAAGTCAGCTATGCTCTTTTGTAGTACCAGTATGTCGACACACACATCTCTTTGAGTT 650
Db 684 GCAATCTGACACCGTAACTTCGATATCTACTATCCAGACTCCAGCATCATCTTTGAGTT 743
Qy 651 CTTTATTCAAAATGATCAGTGGCAGAGATGGACACCACTGACAGTGGTAAACT 710
Db 744 TTTCGTCAGATGACCACTGGCCAGCCCAATGCAGATGACTC---CAGGTGATGAAGAC 800
Qy 711 TACAGACAATGGAGATGGGGCTCTCATTTCTGTAATGCTGAATCAGGCACCAACATACT 770
Db 801 CACAGAGAA---AGGATGGGAATTCACAGTGTGGAGCTTAATTCAGAGGCAATTAATGTCCT 857

Qy 771 CTACTGGAGAACTACAGGCATCCCTTATGGGTTCATAGCGGTCTCAAGCCTGTGCTGGTAAA 830
Db 858 CTATTGGAGAACACACAGCCTCTCTCAGTATGGACCAAGTACCACAGCCTGTGCTGCTGAG 917
Qy 831 AAATATCAATTTGAAGGGTGGCGGTACACATCACAATGTTTCTCTTGCAGGCCAGGCAC 890
Db 918 AAACATTTCCCATTAACAGGGGTGGCGGTACACATTCAGATGCTTCCCTCTCAACCTGGGAC 977
Qy 891 ATTTCAGCAAAACACAGGTTTCATTCACCTGTCAGAGTGTGTCCAGAAACACACATTTCTGA 950
Db 978 GTATCCAGACACAGCAGGCTCTCTTCTGCAAACTTTGCCAGGCCAATCTTTATTCAAA 1037
Qy 951 GAAAGGCCAAAGAAATGTATAAGGTGTA---GACGACTCTCAATTTTCAGGATPCAG 1007
Db 1038 TAAAGGAGAACTCTTCCACACAGTGTGACCTGACAAATACTCAGAGAAAGGATCTTC 1097
Qy 1008 TGAGTGTACAGAGCGCCCTCTCTGTACCACAAAGACATTTTCCAGATCCATCTCCATG 1067
Db 1098 TTTCCTGTAACTGTGCGCCAGCTTGCACAGACAAAGATTTATTTCTACACACACAGCCCTG 1157
Qy 1068 TGATCAAGAAAGAAAGACACAGATTAATGTACAATGTGGATAGAGCCCAAAATCTGCCGGA 1127
Db 1158 CGATGCCAACGAGAGACACACACTCATGTACAAATGGCCAGCCGCAAAATCTGTAGCGA 1217
Qy 1128 GGATCTCACAGATGCTATTTAGATTTGCCCTTCTGTGAGAGAAAGGATTTGCCGCTTGG 1187
Db 1218 GGACCTTGAGGGGCGAGTGAAGCTGCCCTGCTGTGTGAAGACCCACTTCCAGTCCGCTG 1277
Qy 1188 CAACCTTGGATTTTATAACAATGGATCATCTTTCTTGGCCTCCCTCTCTCTCTGGAACAT 1247
Db 1278 CAACCCAGGCTTTCTCAAAACCAACACAGCACCTTGCAGCCCTGCCCATATGGTTCTCTA 1337
Qy 1248 TTCAGATGGAACCAAGAAATGTAGACCATGTCCAGGAGAGGAGCCCTGCACTTTGGGTT 1307
Db 1338 CTCCAATGS---CTCAGACTGTACCCGCTGCCCTGCGGACTGAACCTGCTGTGGAT 1394
Qy 1308 TGAATATAAATGGTGGATGCTCTCTCTGGCAACATGAAACCTTCTGCTGCTCAATGTTGG 1367
Db 1395 TGAATATAAATGGTGGACACGCTGCCCAACAAATGAAAGGAGCCGTTCTCAGTGGGAT 1454
Qy 1368 GAATTTCAAAGTCGGATGGAATGAATGTTGGAGCTGGTGGAGATCATATCCACAGTGG 1427
Db 1455 CAATCTCAGATACAGGCGATGACAGGCTGGGAGGTGGCTGTGATCACAATTTACACAGC 1514
Qy 1428 GCTGGAGGTTCTGACAATGATTAACCTTAACTTAACTATCCAGGATTTAAAC 1487
Db 1515 TGCTGGAGCTCAGACAAATGACTTTCATGATTTCTACTCTGCTGTGTGCCAGGATTTAGACC 1574
Qy 1488 ACCAATCTATGACTGGAGCCAC---GGGTCTGAACCTAGGAAGAAATTAACATTTGCTTT 1544
Db 1575 TCCGAGTCCGTTGGCAGACACAGAGAAATGAAGGTGGCCAGATTCACATTTGCTTT 1634
Qy 1545 TGAGACCTCTCTTCAGCTGACTGTGTTTGTACTTTCATGTTGGATATTAATAGAAAAG 1604
Db 1635 TGAGACCTCTCTCTGTCGAACTGTGAGCTTACTTCTATGTTGGTGGTGTCAATTTAGGAC 1694
Qy 1605 TACAAATGTGGTAGAATCTGGGGTGGAAACCAAGGAAAGAAAGCTTACACCATATCAT 1664
Db 1695 CAACACTCTCTGGAGAGCTGGAAAGGTTCCAAAGGCAAAAGCTCTATACCTACATCAT 1754
Qy 1665 CTTTCAAGATGCAACTTTTACATTTTACATGGGCAATTCAGAGAACTTAATCAGGGTCAAGA 1724
Db 1755 TGAGGAGAACATCTACAGAGCTTCACTTGGGCTTCCAGAGGAGCCACTTTTTCATGAGGC 1814
Qy 1725 TAATAGAGCTTTCATCAATGACATGGTGAAGATTTATTTCTATCAGAGCCACTTAATGCACT 1784
Db 1815 AAGCAGGAAGTACACCAATGACGTTGCCAAGATCTACTCCATCAATGTACCAATGTTAT 1874
Qy 1785 TGATGGGTGGGCTCTCATGCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1844
Db 1875 GAATGGGTGGGCTCTCTACTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1934
Qy 1845 GTCTGTCCCTTCCCTCCAGGCCACTACATTTGAGAAAGAAACCAACCACTGCAAGGAATG 1904

Db 1935 CTGACCTCTGTGCTGCTGTTACTATATTGACCGAGATTGAGGAACCTGCCACTCTG 1994
QY 1905 TCCACCTGACACTACCTGCTCCATACATGAGTCTTATGGCAAGAGGCTGTATTCATG 1964
Db 1995 CCCCCCTAACACAATCTGTAAGACCCACAGCCCTTATGGTGTCCAGGCCCTGTGCCCCG 2054
QY 1965 CGGGCTGGGAGTAAACAAATCAGGACCAATTCGGTTTGTCTATAGTCACTGCTTTTCTA 2024
Db 2055 TGGTCAGGAGCAACAAGAACCAAGATCCACTCTCTGTGCTCAATGATTGCACCTTCTC 2114
QY 2025 CCATGAAAGAAATCAGATTTTGGACATGACTTTAGCAACTCAGCAGTGTGGGCTC 2084
Db 2115 ACACAACACTCCACAGGACTTTCAACTACAACCTCTCCGCTTTGGCAAAACACCGTCAC 2174
QY 2085 ATTAATGAATGGCCCGCTTCACTTCCAAAGGAACAAATATCTTCCATTTCTCAATAT 2144
Db 2175 TCTGTGGAGGCCAAGCTTCACTTCCAAAGGGTTGAATATCTTCCATCCTTTACCT 2234
QY 2145 CAGTTATGTGGGCATGAGGGGAAGATGGCTCTCTGTACCAACAATATAACAGACTT 2204
Db 2235 CAGTCTCTGTGAACACAGGTAGGAAATGTCTGTGACCGCAATGTCTACTGACCT 2294
QY 2205 TACAGTAAAGAAATAGTGCAGGCTCAGATGATTACACAAATTTGGTAGGGCATTTGT 2264
Db 2295 CCGATTCTGGGGTGAGTCAGG-----TTCTCCAAATCTATCACAGCCTACGT 2345
QY 2265 ATGCCAGTCAACAATTTCTTCTGAAAGTAAAGGGTTTCCGAGCAGCCTTATCATCACA 2324
Db 2346 CTGCCAGGAGTATCATCCCCCAGAGTGACAGGCTACAGGCCGGGGTTTCTCTACA 2405
QY 2325 ATCCATCTTCTGGCAGATACATTCATAGAGTACAGTGTGAACACACATTTGAAAAATAT 2384
Db 2406 GCCTGTGAGCTTGTCTGAGCTTATTTGGGTGACAAACAGATATGACTCTGGATGAAAT 2465
QY 2385 TAATATAAAGAGATATGTTCCACTTCCACAAAGCTTGGGAATACCGAGCTGATCTTCT 2444
Db 2466 CACTCCCAAGTGAATTTTCCACTGGAGTCTTGGGAATACCGAGCTGATCTTCT 2525
QY 2445 TTATAAGTCTTCTACAGCAACACATCTTGTATTAATGCGCGATCAACTGCTGTGAATAT 2504
Db 2526 TTATAGTCCAAATGATGTACCCAGCTCTGAGTCTTGGGAGATCAACACCATCCGCGT 2585
QY 2505 GAGGTGATCTTACTTAATCTGGACAGAGTGTATTCAGTCCCCCAGCAAGTGGCCAGC 2564
Db 2586 CAGGTGCACTCCACAGAAAACCTGCTGGAAGTTTGTCTGCCAGGAACGTGCTCAGA 2645
QY 2565 AGTACCTGTGATGGGTGAGTCTTCTTCTGAGAGTGTGAGGAGCCTGCAAGAGAGGATTTT 2624
Db 2646 TGGGACCTGTGATGGTGAATTTCCACTTCCACTTCTGTGGGAGAGCGCGCTGTGCGCGT 2705
QY 2625 GTGTACGGAGCATGACTTCCATGAGATTTGAGGAGCCTGCAAGAGAGGATTTTCAAGAAAC 2684
Db 2706 CTGCTCAGTGGGTGACTACCATGCTATGCTGACGACCTGTGTGGCTGGATCCAGANGAC 2765
QY 2685 CTTGTATGTGGAATGAACCTTAATGGTGCATTAAGGAATTTCTTTCCTGAGAAAAA 2744
Db 2766 TACTTACGTGTGNCAGAGAACCAAGCTATCTCTGTGGTGCATTTCTCTGCTGAGCAGAG 2825
QY 2745 GTTGGCAACCTGTGAAACGGTTGACTTTTGGCTGAGGTGGGAGCGGTGTGGGACCTTT 2804
Db 2826 AGTCACCATCTCGAACCATAGATTTTGGCTGGAAGTGGGCATCTCTGCAAGGACCTTG 2885
QY 2805 TACTGCGGTTTTGCTGGTCTGACCTGCTTACTTCTGGAAGAAAGAAATCAAAAGAAAA 2864
Db 2886 TACTGCCATCTGCTCACCGCTCTGACCTGCTACTTTTGGAAAGAAAGATCAAAACTAGA 2945
QY 2865 GAAGACCATTTGAATCTGTCTCA 2887
Db 2946 GTACAAGTACTCCAAGCTGGTGA 2968

RESULT 15

US-10-176-921-37
; Sequence 37, Application US/10176921
; Publication No. US20030027276A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C288
; CURRENT APPLICATION NUMBER: US/10/176,921
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 37
; LENGTH: 3501
; TYPE: DNA
; ORGANISM: Homo Sapien
; NAME/KEY: unsure
; LOCATION: 2762, 2778
; OTHER INFORMATION: unknown base
US-10-176-921-37

Query Match 26.6%; Score 901; DB 9; Length 3501;
Best Local Similarity 60.0%; Pred. No. 2,5e-241;
Matches 1634; Conservative 0; Mismatches 1062; Indels 27; Gaps 7;

QY 171 GCTTCCTCTTCCAGGAGAAAGATTATCATTGTAATATACGGAATGTGATAGCAGTGG 230
Db 267 GCTTCATGCTGCAAGAGTCTGAGTACCACATATGATACACGGCTGTGACACAGCGG 326
QY 231 CTCACAGTGGAGTGTCCATTCCAAATTTCTGCACTGGAGTGTCTGCGCTGCTGACCC 290
Db 327 TTCCAGTGGAGGTGCGGTGCGCATACCCCGGCGCTGTGCACACAGCTGTCTGACCC 386
QY 291 AGTGAGAGGCAAGAATGCACTTTCTCCTGCTGCTTCTGGAGAGTATCTAGAATGAAGA 350
Db 387 CGTCAAGGCGACGGAGTGTCTCTTCTGCAACGCGGGGAGTTTCTGGATATGAAGA 446
QY 351 CCAGGTATGACAGTAAAGTGTGAGGACACCTATTCTTTGGGAGTGGCATCAAAATTTGA 410
Db 447 CCAGTCTATGAAGCCATGCGCTGAGGCGCTACTCCCTCGGCACAGGATTCGGTTTGA 506
QY 411 TGAATGGGATGAATTCGCGGAGGATTTTCTAACATCGCAACATTTCTGACACATCTGGT 470
Db 507 TGAGTGGGATGAGTGCCTCATGGCTTTCGACGCTCTCAGCCAAACATGGAGCTGGATGA 566
QY 471 GGGCCCTCTGACAGCAGCCGACAGCGCTGTAAACACTTCTTCTGGATCCCTGCTGGA 530
Db 567 CAGTGTGCTGAGTCCA---CCGGGAACGTACTTCTGCTCAAGTGGGTTCCTCCGGGCGCA 623
QY 531 CTACATAGAAATCTAATGCTGATGACTGCAAGTGTCTTTGATCTATGCTGTGCAACCTTAA 590
Db 624 CTACATCGCTCCACACGAGCAATGACACGCCACACTGATGATGACCGCTCAACCTGAA 683
QY 591 GAAGTACAGCTATGCTTCTTTGAGTACAGTATGTGCAACAAACATCTTCTTTGAGTT 650
Db 684 GCAATCTGGCACCGTTAACTTTCGAATACTACTTCCAGACTCCAGCATCATCTTTGAGTT 743

Db 2886 TACTGCCATCTGCTCACCGTCTTGACCTGCTACTTTTGAAAAAGAACTCAAAAACCTAGA 2945

QY 2865 GAAGACCATTGTGATCTGTCA 2887

Qy 2003 GAGGACCAATTCATTCTGGTGA 2967
| | | | |
Db 2946 GTACAAGTACTCCAAGCTGGTGA 2968

Search completed: May 12, 2003, 13:07:24
Job time : 541.238 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.
OM nucleic - nucleic search, using sw model
Run on: May 12, 2003, 00:15:42 ; Search time 173 258 Seconds
(without alignments)
6000.498 Million cell updates/sec
Title: US-10-073-333A-1
Perfect score: 3390
Sequence: 1 atgtgttcgcgcgcgcgggg.....ggtataaaaaaaaaaaaaa 3390
Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA : *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq : *
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq : *
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq : *
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq : *
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq : *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq : *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	493.2	14.5	1129	4	US-09-227-357-40
2	43.2	1.3	1120	3	US-09-030-613-1
3	43.2	1.3	1120	4	US-09-451-905-1
4	42	1.2	1209	4	US-09-105-537-21
5	42	1.2	13613	4	US-09-105-537-3
6	42	1.2	38506	3	US-09-320-878-19
7	41.8	1.2	1743	3	US-08-665-259-20
8	41.8	1.2	1743	3	US-08-762-500-20
9	41.8	1.2	1974	3	US-08-762-500-78
10	41.8	1.2	6803	3	US-08-665-259-19
11	41.8	1.2	6803	3	US-08-762-500-19
12	41.2	1.2	2247	2	US-08-524-828-1
13	41.2	1.2	2247	2	US-08-975-114A-1
14	40.8	1.2	16389	4	US-09-741-154-3
15	40.2	1.2	4403765	4	US-09-103-840A-2
16	40.2	1.2	4411529	4	US-09-103-840A-1
17	39.8	1.2	1652	3	US-08-758-662-3
18	39.6	1.2	2834	4	US-09-305-384-6
19	39.6	1.2	6235	4	US-09-305-384-5
20	39.6	1.2	6679	4	US-09-305-384-1
21	39.4	1.2	3507	4	US-08-775-009-36
22	38.8	1.1	9521	4	US-08-972-218-2
23	38.4	1.1	1283	1	US-07-885-970A-17
24	38.4	1.1	1283	1	US-08-298-687A-17
25	38.4	1.1	1283	1	US-08-530-797-18
26	38.4	1.1	1283	1	US-08-298-829-17
27	38.4	1.1	1283	2	US-08-787-335-18

28	38.4	1.1	2371	2	US-08-343-443B-1	Sequence 1, Appli
29	38.2	1.1	319	1	US-07-593-657-14	Sequence 14, Appli
30	38.2	1.1	1224	1	US-08-924-847A-1	Sequence 1, Appli
31	38.2	1.1	1224	1	US-08-924-847A-3	Sequence 3, Appli
32	38.2	1.1	1224	3	US-09-120-052-1	Sequence 1, Appli
33	38.2	1.1	1224	3	US-09-120-052-3	Sequence 3, Appli
34	38.2	1.1	1415	4	US-09-345-214-11	Sequence 11, Appli
35	38.2	1.1	1534	1	US-08-300-903A-6	Sequence 6, Appli
36	38.2	1.1	2008	4	US-09-345-214-12	Sequence 12, Appli
37	38.2	1.1	2491	4	US-09-345-214-5	Sequence 5, Appli
38	38.2	1.1	4348	2	US-08-915-868-1	Sequence 1, Appli
39	38	1.1	501	2	US-08-997-080-107	Sequence 107, App
40	38	1.1	501	2	US-08-997-362-107	Sequence 107, App
41	38	1.1	501	4	US-09-095-855-107	Sequence 107, App
42	38	1.1	501	4	US-09-324-542-107	Sequence 107, App
43	38	1.1	501	4	US-09-205-426-107	Sequence 107, App
44	38	1.1	2007	3	US-08-747-221B-36	Sequence 36, Appl
45	38	1.1	2007	3	US-08-747-221B-38	Sequence 38, Appl

ALIGNMENTS

RESULT 1
US-09-227-357-40
; Sequence 40, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT APPLICATION NUMBER: US/09/227,357
; CURRENT FILING DATE: 1999-01-08
; EARLIER APPLICATION NUMBER: PCT/US98/13684
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: 60/051,926
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,793
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,925
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,929
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,803
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,732
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,931
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,932
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,916
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,930
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,918
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,920
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,733
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,795
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,919
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,928
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/055,722
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,723
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,948
; EARLIER FILING DATE: 1997-08-18

EARLIER APPLICATION NUMBER: 60/055,949
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,953
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,950
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,947
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,964
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/056,360
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,684
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,984
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/055,954
EARLIER FILING DATE: 1997-08-18
EARLIER APPLICATION NUMBER: 60/058,785
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,664
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,660
EARLIER FILING DATE: 1997-09-12
EARLIER APPLICATION NUMBER: 60/058,661
EARLIER FILING DATE: 1997-09-12
NUMBER OF SEQ ID NOS: 672
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 40
LENGTH: 1129
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (1053)
OTHER INFORMATION: n equals a,t,g, or c
US-09-227-357-40

Query Match 14.5%; Score 493.2; DB 4; Length 1129;
Best Local Similarity 98.3%; Pred. No. 2.9e-130;
Matches 509; Conservative 0; Mismatches 8; Indels 1; Gaps 1;
QY 2860 AAAAGAAGACCATTTTGAATCTGTTCACTGAAACCTCAAGATCCCCCAATATATGAA 2919
DB 446 AAAAGAAGACCATTTTGAATCTGTTCACTGAAACCTCAAGATCCCCCAATATATGAA 505
QY 2920 GAGACAGTCTGTAG-CCYTTGAGACTAATGAACAAAGAAACCTGCTCTAGTTTACAGGA 2978
DB 506 GAGACAGTCTGTAGCCCTTTGAGACTAATGAACAAAGAAACCTGCTCTAGTTTACAGGA 565
QY 2979 CCATATTTTAGGCTCTGCTCTACATCTGTACATTTGGTGATCTCACAGAGAGGCCAT 3038
DB 566 CCATATTTTAGGCTCTGCTCTACATCTGTACATTTGGTGATCTCACAGAGAGGCCAT 625
QY 3039 GCCGCTGAAAAGGAGGAGATTTGAACATTTGATTCGCTTATACATGTCACAGTACCT 3098
DB 626 GCCGCTGAAAAGGAGGAGATTTGAACATTTGATTCGCTTATACATGTCACAGTACCT 685
QY 3099 TGCCAAATAAGGAAGAAACCAATGATTTGGGCTCAACTGAAGATGAAGCTCAACTCAGGA 3158
DB 686 TGCCAAATAAGGAAGAAACCAATGATTTGGGCTCAACTGAAGATGAAGCTCAACTCAGGA 745
QY 3159 AGAGATTATCTGTATATACATAAATCACTGAAACCAAGTTTAAAGCCCACTGCTGC 3218
DB 746 AGAGATTATCTGTATATACATAAATCACTGAAACCAAGTTTAAAGCCCACTGCTGC 805
QY 3219 TGATGATGCCATATATTAATGGTAACTTTTATCTTATGATGCTCATATAACAGT 3278
DB 806 TGATGATGCCATATATTAATGGTAACTTTTATCTTATGATGCTCATATAACAGT 865
QY 3279 GTGATTTGGAAGGACACATGTGAGCATATGCAATATGATCCAAATTTATGTTTCTTTGT 3338
DB 866 GTGATTTGGAAGGACACATGTGAGCATATGCAATATGATCCAAATTTATGTTTCTTTGT 925

QY 3339 TTATATTTTGGGAAATTAATATTTTAAAGGTAA 3376
DB 926 TTATATTTTGGGAAATTAATATTTTAAAGGTATA 963
RESULT 2
US-09-030-613-1
Sequence 1, Application US/09030613
Patent No. 6083706
GENERAL INFORMATION:
APPLICANT: Florkiewicz, Robert Z.
APPLICANT: Baird, J. Andrew
TITLE OF INVENTION: INHIBITORS OF LEADERLESS PROTEIN EXPORT
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/09/030,613
APPLICATION NUMBER: US/09/030,613
FILING DATE: 25-FEB-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: No. 6083706tenburg Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 760100.418C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1120 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-030-613-1
Query Match 1.3%; Score 43.2; DB 3; Length 1120;
Best Local Similarity 63.5%; Pred. No. 0.039; Mismatches 38; Indels 0; Gaps 0;
Matches 66; Conservative 0
QY 4 CTGTTCCGCGCGCGGCGGCTACGGGCGAGGGCTGGGGCGCGCGCGCGGAGGCTCC 63
DB 258 CGGGGCGCTGCCCCGAGCGGGTCTGGAGCGCGGGCGGGGCGGGGCGGCGGCTCC 317
QY 64 CGCGCGCGCGCTGCGCGCGCGCTGGAGCGCGCGCGCTGATTTGCTG 107
DB 318 CGCGCGCGCTCCAGCGGCTCGGGGATCCGGGCGCGCGCGCGCGCGCG 361
RESULT 3
US-09-451-905-1
Sequence 1, Application US/09451905
Patent No. 6306613
GENERAL INFORMATION:
APPLICANT: Robert Z. Florkiewicz
APPLICANT: Andrew Baird
APPLICANT: Dale E. Warnock
TITLE OF INVENTION: MODULATORS OF LEADERLESS PROTEIN EXPORT
FILE REFERENCE: 200124.402C4
CURRENT APPLICATION NUMBER: US/09/451,905
CURRENT FILING DATE: 1999-12-01
NUMBER OF SEQ ID NOS: 48

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1120
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-451-905-1

Query Match      1.3%; Score 43.2; DB 4; Length 1120;
Best Local Similarity 63.5%; Pred. No. 0.039;
Matches 66; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 4  CTGTTCCGGCCCGGGGGCGGTACGGGCGAGGCGCTGGGGCGCGCGCGGAGGCTCCC 63
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 258 CGGGCGCGTGCCTCCGAGGCGGTGAGGCGCGCGCGCGCGCGGCGGCGGCGCTCCC 317
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 64  CGCGCGGCGCTCGCCGCCCTGAGCCCGCCCTGGATTGCTG 107
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 318 CGCGCGGCTCGAGGCGTGGGGATCCCGCGCGGCGCGCGCGAG 361
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 4
US-09-105-537-21
; Sequence 21, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438U51
; CURRENT APPLICATION NUMBER: US/09/105.537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 1209
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-21

Query Match      1.2%; Score 42; DB 4; Length 1209;
Best Local Similarity 62.3%; Pred. No. 0.089;
Matches 66; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 28  CGGGCAGGGCTGGGGCGCGCGGAGGCTCCCGCGCGCGCGCGCTCGCGCCCTGG 87
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 821 CGGGCGCGGTGGAGAGACCTCCGCTACGACCCCGCGGTGAGCTCGAGCGCCGGG 880
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 88  AGCCCCCGCTGGATTGCTGGGCGCTCGCGGCTGCCAGGCGG 133
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 881 TGGTCCGGGGAGACGAGGCTGGCGGCGCGCGCTCGCGCGCGG 926
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 5
US-09-105-537-3
; Sequence 3, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438U51
; CURRENT APPLICATION NUMBER: US/09/105.537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 13613
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
```

US-09-105-537-3

```
Query Match      1.2%; Score 42; DB 4; Length 13613;
Best Local Similarity 62.3%; Pred. No. 0.44;
Matches 66; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
```

```
QY 28  CGGGCAGGGCTGGGGCGCGCGGAGGCTCCCGCGCGCGCGGCTCGCGCCCTGG 87
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1626 CGGCGCGCGGTGGAGAGACCTCCGCTACGACCCCGCGGTGAGCTCGAGCGCCGG 1685
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 88  AGCCCCCGCTGGATTGCTGGGCGCTCGCGCGGCTGCCAGGCGG 133
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1686 TGGTCCGGGGAGACGAGGCTGGCGGCGCGGCGGCTGCCGCGCGG 1731
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

RESULT 6

```
US-09-320-878-19
; Sequence 19, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320.878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141.908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073.538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846.247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-320-878-19
```

```
Query Match      1.2%; Score 42; DB 3; Length 38506;
Best Local Similarity 62.3%; Pred. No. 0.88;
Matches 66; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
```

```
QY 28  CGGGCAGGGCTGGGGCGCGCGGAGGCTCCCGCGCGCGGCTCGCGCCCTGG 87
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 35623 CGGCGCGCGGTGGAGAGACCTCCGCTACGACCCCGCGGTGAGCTCGAGCGCCCGG 35682
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 88  AGCCCCCGCTGGATTGCTGGGCGCTCGCGGCTGCCAGGCGG 133
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 35683 TGGTCCGGGGAGACGAGGCTGGCGGCGCGGCGGCTGCCGCGCGG 35728
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

RESULT 7

```
US-08-665-259-20
; Sequence 20, Application US/08665259
; Patent No. 6028173
; GENERAL INFORMATION:
; APPLICANT: Landes, Gregory M.
; APPLICANT: Burn, Timothy C.
; APPLICANT: Connors, Timothy D.
; APPLICANT: Dackowski, William R.
; APPLICANT: Van Raay, Terence J.
; APPLICANT: Klinger, Katherine W.
```

;; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
;; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
;; NUMBER OF SEQUENCES: 73
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: GENZYME CORPORATION
;; STREET: One Mountain Road
;; CITY: Framingham
;; STATE: Massachusetts
;; COUNTRY: United States of America
;; ZIP: 01701
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/665,259
;; FILING DATE: 17-JUN-1996
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/665,259
;; FILING DATE: 17-JUN-1996
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US96/10469
;; FILING DATE: 17-JUN-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Dugan, Deborah A.
;; REGISTRATION NUMBER: 37,315
;; REFERENCE/DOCKET NUMBER: IG5-9.1
;; TELEPHONE: (508) 872-8400
;; TELEFAX: (508) 872-5415
;; INFORMATION FOR SEQ ID NO: 20:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1743 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1..1740
US-08-665-259-20

Query Match 1.2%; Score 41.8; DB 3; Length 1743;
Best Local Similarity 59.8%; Pred. No. 0.13;
Matches 70; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 6 GTTCCGGCGCGCGCGCTGACGGGCGAGGGCTGGGGGGCGCGGAGGCTCCCGG 65
Db 1505 GTAGCGCGCTGTGGTGCGCCCGGGGATGCGGCTGCGGCGCTGCTCCCGG 1564

Qy 66 CCGCGGGCGCTCGCGCCCTGGAGCCCGCCCTGGATTGCTGCTGGGGCGCTCGCGG 122
Db 1565 GCCGCGCTACTCTGCTGCTGGGGGGCGGCTGGAGCCGCGCTGGGGGGCGGGG 1621

RESULT 8
US-08-762-500-20
; Sequence 20, Application US/08762500
; Patent No. 6030806
; GENERAL INFORMATION:
; APPLICANT: Landes, Gregory M.
; APPLICANT: Burn, Timothy C.
; APPLICANT: Connors, Timothy D.
; APPLICANT: Dackowski, William R.
; APPLICANT: Van Raay, Terence J.
; APPLICANT: Klingner, Katherine W.
; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
; NUMBER OF SEQUENCES: 83
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: One Mountain Road
; CITY: Framingham
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/762,500
;; FILING DATE: 09-DEC-1996
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/665,259
;; FILING DATE: 17-JUN-1996
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US96/10469
;; FILING DATE: 17-JUN-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Dugan, Deborah A.
;; REGISTRATION NUMBER: 37,315
;; REFERENCE/DOCKET NUMBER: IG5-9.3
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (508) 872-8400
;; TELEFAX: (508) 872-5415
;; INFORMATION FOR SEQ ID NO: 20:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1743 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 1..1740
US-08-762-500-20

Query Match 1.2%; Score 41.8; DB 3; Length 1743;
Best Local Similarity 59.8%; Pred. No. 0.13;
Matches 70; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 6 GTTCCGGCGCGCGCGCTGACGGGCGAGGGCTGGGGGGCGCGGAGGCTCCCGG 65
Db 1505 GTAGCGCGCTGTGGTGCGCCCGGGGATGCGGCTGCGGCGCTGCTCCCGG 1564

Qy 66 CCGCGGGCGCTCGCGCCCTGGAGCCCGCCCTGGATTGCTGCTGGGGCGCTCGCGG 122
Db 1565 GCCGCGCTACTCTGCTGCTGGGGGGCGGCTGGAGCCGCGCTGGGGGGCGGGG 1621

RESULT 9
US-08-762-500-78
; Sequence 78, Application US/08762500
; Patent No. 6030806
; GENERAL INFORMATION:
; APPLICANT: Landes, Gregory M.
; APPLICANT: Burn, Timothy C.
; APPLICANT: Connors, Timothy D.
; APPLICANT: Dackowski, William R.
; APPLICANT: Van Raay, Terence J.
; APPLICANT: Klingner, Katherine W.
; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
; NUMBER OF SEQUENCES: 83
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: One Mountain Road
; CITY: Framingham
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/762,500
FILING DATE: 09-DEC-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/665,259
FILING DATE: 17-JUN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/10469
FILING DATE: 17-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Dugan, Deborah A.
REGISTRATION NUMBER: 37,315
REFERENCE/DOCKET NUMBER: IG5-9.3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 872-8400
TELEFAX: (508) 872-5415
INFORMATION FOR SEQ ID NO: 78:
SEQUENCE CHARACTERISTICS:
LENGTH: 1974 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-762-500-78

Query Match 1.2%; Score 41.8; DB 3; Length 1974;
Best Local Similarity 59.8%; Pred. No. 0.14; Mismatches 0; Gaps 0;
Matches 70; Conservative 0;
QY 6 GTTCCGCGCCGGGCGGTACGGGCGAGGGCTGGGGCGCGCGGAGGCTCCCG 65
|| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1714 GTAGCGCGCTGTGGTGCCCGCGGGATGCGCGCTGCGCGCTGCTCCCG 1773
QY 66 CCGCGGCGCTCGCGCCCTGGAGCCCGCCCTGGATTTCGTGCTGGCGCTGCCCG 122
|| || || || || || || || || || || || || || || || || || || || ||
Db 1774 GCGCGCGCTACCTCTCTGCTGGGGGCGCGCTGGAGCCGCGCTGGGGCGCGGGG 1830

RESULT 10
US-08-665-259-19
Sequence 19, Application US/08665259
Patent No. 6028173
GENERAL INFORMATION:
APPLICANT: Landes, Gregory M.
APPLICANT: Burn, Timothy C.
APPLICANT: Connors, Timothy D.
APPLICANT: Dackowski, William R.
APPLICANT: Van Raay, Terence J.
APPLICANT: Klingler, Katherine W.
TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
METHODS OF MAKING AND USING SAME
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENZYME CORPORATION
STREET: One Mountain Road
CITY: Framingham
STATE: Massachusetts
COUNTRY: United States of America
ZIP: 01701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/665,259
FILING DATE: 17-JUN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Dugan, Deborah A.
REGISTRATION NUMBER: 37,315
REFERENCE/DOCKET NUMBER: IG5-9.1
TELECOMMUNICATION INFORMATION:

TELEPHONE: (508) 872-8400
TELEFAX: (508) 872-5415
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 6803 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-665-259-19
Query Match 1.2%; Score 41.8; DB 3; Length 6803;
Best Local Similarity 59.8%; Pred. No. 0.32; Mismatches 70; Conservative 0; Gaps 0;
Matches 70; Conservative 0;
QY 6 GTTCCGCGCCGGGCGGTACGGGCGAGGGCTGGGGCGCGCGGAGGCTCCCG 65
|| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5298 GTAGCGCGCTGTGGTGCCCGCGGGATGCGCGCTGCGCGCTGCTCCCG 5357
QY 66 CCGCGGCGCTCGCGCCCTGGAGCCCGCCCTGGATTTCGTGCTGGCGCTGCCCG 122
|| || || || || || || || || || || || || || || || || || || || ||
Db 5358 GCGCGCGCTACCTCTCTGCTGGGGGCGCGCTGGAGCCGCGCTGGGGCGCGGGG 5414

RESULT 11
US-08-762-500-19
Sequence 19, Application US/08762500
Patent No. 6030806
GENERAL INFORMATION:
APPLICANT: Landes, Gregory M.
APPLICANT: Burn, Timothy C.
APPLICANT: Connors, Timothy D.
APPLICANT: Dackowski, William R.
APPLICANT: Van Raay, Terence J.
APPLICANT: Klingler, Katherine W.
TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
METHODS OF MAKING AND USING SAME
NUMBER OF SEQUENCES: 83
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENZYME CORPORATION
STREET: One Mountain Road
CITY: Framingham
STATE: Massachusetts
COUNTRY: United States of America
ZIP: 01701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/762,500
FILING DATE: 09-DEC-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/665,259
FILING DATE: 17-JUN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/10469
FILING DATE: 17-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Dugan, Deborah A.
REGISTRATION NUMBER: 37,315
REFERENCE/DOCKET NUMBER: IG5-9.3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 872-8400
TELEFAX: (508) 872-5415
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 6803 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

```
MOLECULE TYPE: DNA (genomic)
US-08-762-500-19

Query Match
Best Local Similarity 1.2%; Score 41.8; DB 3; Length 6803;
Matches 70; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 6 GTTCGCGCCCGGGCGGCTACGGGCGAGGGCTGGGGCGCGCGCGAGGCTCCCGG 65
  || ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5298 GTAGCGCGTGTGGTGGCGCGGGGATGCGCGCTGCGCGCTGCTCCCGG 5357

QY 66 CGCGGGCGCTCGCGCCCTGGAGCCCCCGCTGGATTGCTGCGCGCTCGCGGG 122
  || || || || || || || || || || || || || || || || || || ||
Db 5358 GCGCGCGCTACCTCCTGCTGGGGGGCGGCGCTGGAGCCGCGCTGGGGCGGGGG 5414

RESULT 12
US-08-524-828-1
; Sequence 1, Application US/08524828
; Patent No. 5874271
; GENERAL INFORMATION:
; APPLICANT: Atsushi NISHIKAWA et al.
; TITLE OF INVENTION: HUMAN GLYCOSYLTRANSFERASE GENE,
; COMPOUNDS AND METHOD FOR INHIBITING CANCEROUS METASTASIS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/524,828
; FILING DATE: September 7, 1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/107,173
; FILING DATE: August 17, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: James H. Meadows
; REGISTRATION NUMBER: 33,965
; REFERENCE/DOCKET NUMBER: JTF/001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2247 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
US-08-524-828-1

Query Match
Best Local Similarity 1.2%; Score 41.2; DB 2; Length 2247;
Matches 64; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 9 CGCGCGCGCGGGCGGCTACGGGCGAGGGCTGGGGCGCGCGCGAGGCTCCCGCGG 68
  || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 33 CTGCGCGCGGACCGCGGGATGGGGCGAGAGGCTGGCGGCGAGCCAGCATCTCCCGCGG 92

QY 69 CGGGCGCTCGCGCGCTGGAGCCCCCGCTGGATTGCTGCTG 110
  || || || || || || || || || || || || || || || || || || ||
Db 93 GGGACCGCGGGCGCGAGCGCGCGCGCGCTGCTGCGG 134

US-08-524-828-1

Query Match
Best Local Similarity 1.2%; Score 41.2; DB 2; Length 2247;
Matches 64; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 9 CGCGCGCGCGGGCGGCTACGGGCGAGGGCTGGGGCGCGCGCGAGGCTCCCGCGG 68
  || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 33 CTGCGCGCGGACCGCGGGATGGGGCGAGAGGCTGGCGGCGAGCCAGCATCTCCCGCGG 92

QY 69 CGGGCGCTCGCGCGCTGGAGCCCCCGCTGGATTGCTGCTG 110
  || || || || || || || || || || || || || || || || || || ||
Db 93 GGGACCGCGGGCGCGAGCGCGCGCGCGCTGCTGCGG 134
```

```
RESULT 13
US-08-975-114A-1
; Sequence 1, Application US/08975114A
; Patent No. 5876714
; GENERAL INFORMATION:
; APPLICANT: Atsushi NISHIKAWA et al.
; TITLE OF INVENTION: HUMAN GLYCOSYLTRANSFERASE GENE,
; COMPOUNDS AND METHOD FOR INHIBITING CANCEROUS METASTASIS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., Suite 800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/975,114A
; FILING DATE: No. 5876714ember 20, 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/107,173
; FILING DATE: August 17, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee Cheng
; REGISTRATION NUMBER: 40,949
; REFERENCE/DOCKET NUMBER: 1-F3439DIV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2247 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
US-08-975-114A-1

Query Match
Best Local Similarity 1.2%; Score 41.2; DB 2; Length 2247;
Matches 64; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 9 CGCGCGCGCGGGCGGCTACGGGCGAGGGCTGGGGCGCGCGCGAGGCTCCCGCGG 68
  || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 33 CTGCGCGCGGACCGCGGGATGGGGCGAGAGGCTGGCGGCGAGCCAGCATCTCCCGCGG 92

QY 69 CGGGCGCTCGCGCGCTGGAGCCCCCGCTGGATTGCTGCTG 110
  || || || || || || || || || || || || || || || || || || ||
Db 93 GGGACCGCGGGCGCGAGCGCGCGCGCGCTGCTGCGG 134

RESULT 14
US-09-741-154-3/c
; Sequence 3, Application US/09741154
; Patent No. 6437110
; GENERAL INFORMATION:
; APPLICANT: BEASLEY, Ellen M. et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; THEREOF
; FILE REFERENCE: CL001061
; CURRENT APPLICATION NUMBER: US/09/741,154
; CURRENT FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
```

Search completed: May 12, 2003, 11:08:26
Job time : 11856.3 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 05:40:13 ; Search time 546.762 Seconds
(without alignments)
8082.236 Million cell updates/sec

Title: US-10-073-333A-3
Perfect score: 3556
Sequence: 1 atgctgttcgcgcgcgcggg.....ggtataaaaaaaaaaaaaa 3556

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 783854 seqs, 621352466 residues

Total number of hits satisfying chosen parameters: 1567708

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_NA:

```
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3556	100.0	3556	9	US-10-140-164-3
2	3214	90.4	3390	9	US-10-140-164-1
3	1499.6	42.2	1737	9	US-10-002-050-19
4	1499.6	42.2	1737	9	US-10-002-304-19
5	1499.6	42.2	1737	12	US-10-003-152-19
6	1182.6	33.3	1508	9	US-10-002-050-9
7	1182.6	33.3	1508	9	US-10-002-304-9
8	1182.6	33.3	1508	12	US-10-003-152-9
9	987.8	27.8	3501	9	US-10-028-072-37
10	987.8	27.8	3501	9	US-10-121-049-37
11	987.8	27.8	3501	9	US-10-123-904-37
12	987.8	27.8	3501	9	US-10-140-470-37
13	987.8	27.8	3501	9	US-10-175-746-37
14	987.8	27.8	3501	9	US-10-176-918-37
15	987.8	27.8	3501	9	US-10-176-921-37
16	987.8	27.8	3501	9	US-10-137-865-37
17	987.8	27.8	3501	9	US-10-140-474-37
18	987.8	27.8	3501	9	US-10-142-431-37
19	987.8	27.8	3501	9	US-10-143-114-37

```
20 987.8 27.8 3501 9 US-10-140-002-37 Sequence 37, Appl
21 987.8 27.8 3501 9 US-10-142-419-37 Sequence 37, Appl
22 987.8 27.8 3501 9 US-10-123-262-37 Sequence 37, Appl
23 987.8 27.8 3501 9 US-10-142-423-37 Sequence 37, Appl
24 987.8 27.8 3501 9 US-10-121-050-37 Sequence 37, Appl
25 987.8 27.8 3501 9 US-10-141-755-37 Sequence 37, Appl
26 987.8 27.8 3501 9 US-10-143-032-37 Sequence 37, Appl
27 987.8 27.8 3501 9 US-10-123-108-37 Sequence 37, Appl
28 987.8 27.8 3501 9 US-10-123-236-37 Sequence 37, Appl
29 987.8 27.8 3501 9 US-10-123-261-37 Sequence 37, Appl
30 987.8 27.8 3501 9 US-10-140-921-37 Sequence 37, Appl
31 987.8 27.8 3501 9 US-10-140-928-37 Sequence 37, Appl
32 987.8 27.8 3501 9 US-10-121-045-37 Sequence 37, Appl
33 987.8 27.8 3501 9 US-10-123-292-37 Sequence 37, Appl
34 987.8 27.8 3501 9 US-10-123-903-37 Sequence 37, Appl
35 987.8 27.8 3501 9 US-10-124-819-37 Sequence 37, Appl
36 987.8 27.8 3501 9 US-10-124-822-37 Sequence 37, Appl
37 987.8 27.8 3501 9 US-10-140-925-37 Sequence 37, Appl
38 987.8 27.8 3501 9 US-10-160-498-37 Sequence 37, Appl
39 987.8 27.8 3501 9 US-10-121-041-37 Sequence 37, Appl
40 987.8 27.8 3501 9 US-10-121-043-37 Sequence 37, Appl
41 987.8 27.8 3501 9 US-10-121-047-37 Sequence 37, Appl
42 987.8 27.8 3501 9 US-10-123-215-37 Sequence 37, Appl
43 987.8 27.8 3501 9 US-10-123-902-37 Sequence 37, Appl
44 987.8 27.8 3501 9 US-10-123-908-37 Sequence 37, Appl
45 987.8 27.8 3501 9 US-10-123-909-37 Sequence 37, Appl
```

ALIGNMENTS

```
RESULT 1
US-10-140-164-3
; Sequence 3, Application US/10140164
; Publication No. US2003007236A1
; GENERAL INFORMATION:
; APPLICANT: Baker et al.
; TITLE OF INVENTION: Human Tumor Necrosis Factor Receptor TR16
; FILE REFERENCE: PF514C1
; CURRENT APPLICATION NUMBER: US/10/140,164
; CURRENT FILING DATE: 2002-05-08
; PRIOR APPLICATION NUMBER: 09/637,856
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/148,348
; PRIOR FILING DATE: 1999-08-12
; PRIOR APPLICATION NUMBER: 60/148,683
; PRIOR FILING DATE: 1999-08-13
; PRIOR APPLICATION NUMBER: 60/148,870
; PRIOR FILING DATE: 1999-08-13
; PRIOR APPLICATION NUMBER: 60/148,758
; PRIOR FILING DATE: 1999-08-16
; PRIOR APPLICATION NUMBER: 60/149,181
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: 60/149,453
; PRIOR FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: 60/149,498
; PRIOR FILING DATE: 1999-08-19
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 3556
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3084)
US-10-140-164-3
```

```
Query Match 100.0%; Score 3556; DB 9; Length 3556;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3556; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGCTGTTCCGCGCGCGCGCGGTACGGGGCTGGGGCGCGCGAGGCT 60
```

|||||
Db 1 ATGCTGTTCCGCGCGGGGCGCGGTACGGGCAAGGGCTGGGGCGCGCGGAGGCT 60
Qy 61 CCCCGCGCGGGCGTCCGCCCTCGAGCCCGCGCTGGATTTGCTGTGGCGCTGCC 120
Db 61 CCCCGCGCGGGCGTCCGCCCTCGAGCCCGCGCTGGATTTGCTGTGGCGCTGCC 120
Qy 121 GCGTCCAGGGCGGCTGGGCTGGGACCTGCGCTCTCCAGCGCGCGCGCTTCCCT 180
Db 121 GCGTCCAGGGCGGCTGGGCTGGGACCTGCGCTCTCCAGCGCGCGCGCTTCCCT 180
Qy 181 TCCAGGAGAAAGATTATCACTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGTGG 240
Db 181 TCCAGGAGAAAGATTATCACTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGTGG 240
Qy 241 AGATTTGCCATTCCAAATTTCTGACGTGACCTGCTGCGCTGCCCTGACCCAGTGAGAGGC 300
Db 241 AGATTTGCCATTCCAAATTTCTGACGTGACCTGCTGCGCTGCCCTGACCCAGTGAGAGGC 300
Qy 301 AAAGAATGCATTTCTCCTGCTCTCTGGAGAGTATCTAGAAATGAAGAACAGATATGC 360
Db 301 AAAGAATGCATTTCTCCTGCTCTCTGGAGAGTATCTAGAAATGAAGAACAGATATGC 360
Qy 361 AGTAAGTGTGTGAAGGCACCTATTCTTGGCAGTGGCATCAAAATTTGATGAATGGGAT 420
Db 361 AGTAAGTGTGTGAAGGCACCTATTCTTGGCAGTGGCATCAAAATTTGATGAATGGGAT 420
Qy 421 GAATTTCCGCGGAGGATTTCTTAACATCCCAACATTCATGGACACTGTGTGGCGCTTCT 480
Db 421 GAATTTCCGCGGAGGATTTCTTAACATCCCAACATTCATGGACACTGTGTGGCGCTTCT 480
Qy 481 GACAGAGCCAGAGCGGTGAACAACTCTTTTGATGCCCTCGTGGAACTACATAGAA 540
Db 481 GACAGAGCCAGAGCGGTGAACAACTCTTTTGATGCCCTCGTGGAACTACATAGAA 540
Qy 541 TCTAATCGTGATGACTGCACGGTGCTTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGC 600
Db 541 TCTAATCGTGATGACTGCACGGTGCTTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGC 600
Qy 601 TATGTCTCTTTGACTACCAGTATGTCGACAAACAACTCTTCTTTGAGTCTTTTATTCAA 660
Db 601 TATGTCTCTTTGACTACCAGTATGTCGACAAACAACTCTTCTTTGAGTCTTTTATTCAA 660
Qy 661 AATGATCAGTCCAGAGATGACACACCACTGACAGTGGGTAAAATTTACAGACAA 720
Db 661 AATGATCAGTCCAGAGATGACACACCACTGACAGTGGGTAAAATTTACAGACAA 720
Qy 721 GCAGAAATGGGCTCTCATTTCTGTAATGCTGAAATCAGGCACAAACATCTCTACTGGAGA 780
Db 721 GCAGAAATGGGCTCTCATTTCTGTAATGCTGAAATCAGGCACAAACATCTCTACTGGAGA 780
Qy 781 ACTACAGGCATCTTTATGGGTTCTAAGGCGGTCAAGCCTGTGCTGGTAAAAATATCACA 840
Db 781 ACTACAGGCATCTTTATGGGTTCTAAGGCGGTCAAGCCTGTGCTGGTAAAAATATCACA 840
Qy 841 ATTGAAGGGGTGGGTACACATCAGAAATGTTTCCCTTGAAGCCAGGCACATTCAGCAAC 900
Db 841 ATTGAAGGGGTGGGTACACATCAGAAATGTTTCCCTTGAAGCCAGGCACATTCAGCAAC 900
Qy 901 AAACAGGTTCAATCAACTGCCAGGTGCTCCAGAAACACCTATTCTCAGAAAGGAGCC 960
Db 901 AAACAGGTTCAATCAACTGCCAGGTGCTCCAGAAACACCTATTCTCAGAAAGGAGCC 960
Qy 961 AAAGAATGTATAGGTGTAAAGCAGCTCTCAATTTTTCAGGATCCAGTGGTGTACAGAG 1020
Db 961 AAAGAATGTATAGGTGTAAAGCAGCTCTCAATTTTTCAGGATCCAGTGGTGTACAGAG 1020
Qy 1021 CGCCCTCCCTGTACCAACAAAGACTATTTCCAGATCCATCTACTGTGATGAAGAAGA 1080
Db 1021 CGCCCTCCCTGTACCAACAAAGACTATTTCCAGATCCATCTACTGTGATGAAGAAGA 1080
Qy 1081 AAGACACAGATTAATGTACAAAGTGGATAGAGCCCAAAATCTCGCGGAGGATCTCACAGAT 1140
|||||

Db 1081 AAGACACAGATTAATGTACAAAGTGGATACAGCCCAAAATCTGCGGGAGGATCTCACAGAT 1140
Qy 1141 GCTATTAGATTGCCCTTCTGGAGAGAAGAGGATTTGTCGCCCTTGGCAACCTGGATTT 1200
Db 1141 GCTATTAGATTGCCCTTCTGGAGAGAAGAGGATTTGTCGCCCTTGGCAACCTGGATTT 1200
Qy 1201 TATACAAATGGATCATCTTCTTGCCATCCCCTGCTCCTCGGAACATTTTCAGATGGAACC 1260
Db 1201 TATACAAATGGATCATCTTCTTGCCATCCCCTGCTCCTCGGAACATTTTCAGATGGAACC 1260
Qy 1261 AAAGAATGTAGACCATGTCCAGCAGGAAGAGCCTGCACCTTGGCTTTGAAATATAATGG 1320
Db 1261 AAAGAATGTAGACCATGTCCAGCAGGAAGAGCCTGCACCTTGGCTTTGAAATATAATGG 1320
Qy 1321 TGGAAATGCTTCTTCCGCAACATGAATACTTCTGCTTCAATTTGTTGGGAATTCAAAGTGC 1380
Db 1321 TGGAAATGCTTCTTCCGCAACATGAATACTTCTGCTTCAATTTGTTGGGAATTCAAAGTGC 1380
Qy 1381 GATGGAATGAATGGTTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGCTCGAGGTCT 1440
Db 1381 GATGGAATGAATGGTTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGCTCGAGGTCT 1440
Qy 1441 GACAATGATTAACCTGATCTTAAACTTGGCATATCCAGAGATTTAAACCCACCAACATCTATG 1500
Db 1441 GACAATGATTAACCTGATCTTAAACTTGGCATATCCAGAGATTTAAACCCACCAACATCTATG 1500
Qy 1501 ACTGGAGCCACGGTTCGAACTAGGAAGATAACATTTTGTCTTTGAGACCCCTCTGTTCA 1560
Db 1501 ACTGGAGCCACGGTTCGAACTAGGAAGATAACATTTTGTCTTTGAGACCCCTCTGTTCA 1560
Qy 1561 GCTGACTGTGTTTGTACTTTCATGCTGGATATTAAGAAAAGTACAAATCTGGTAGAA 1620
Db 1561 GCTGACTGTGTTTGTACTTTCATGCTGGATATTAAGAAAAGTACAAATCTGGTAGAA 1620
Qy 1621 TCGTGGGTGGAAACAAAGAAACAAAGCTTACACCCATATCATCTTCAAGAAATGCAACT 1680
Db 1621 TCGTGGGTGGAAACAAAGAAACAAAGCTTACACCCATATCATCTTCAAGAAATGCAACT 1680
Qy 1681 TTTTACATTTACATGGGCATTTCCAGAGACTAATAGGGTCAAGATAAGAGCGTTCATC 1740
Db 1681 TTTTACATTTACATGGGCATTTCCAGAGACTAATAGGGTCAAGATAAGAGCGTTCATC 1740
Qy 1741 AATGACATGCTCAAGATTTTATTTCTATCAGACCACTAATGCAAGTTCATGGGTGCC 1800
Db 1741 AATGACATGCTCAAGATTTTATTTCTATCAGACCACTAATGCAAGTTCATGGGTGCC 1800
Qy 1801 TCATGCGCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860
Db 1801 TCATGCGCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860
Qy 1861 CCAGGCCACTACATTTGAGAAAGAAACCAACCACTGCAAGGAATGTCACCTGACACCTAC 1920
Db 1861 CCAGGCCACTACATTTGAGAAAGAAACCAACCACTGCAAGGAATGTCACCTGACACCTAC 1920
Qy 1921 CTGTCCATACATCAGTCTATGCAAGAGGCTTGTATTTCCATGCGGGCTGGGAGTAA 1980
Db 1921 CTGTCCATACATCAGTCTATGCAAGAGGCTTGTATTTCCATGCGGGCTGGGAGTAA 1980
Qy 1981 AACAAATCAGGACCATTCGGTTCGTATAGTGACTGCTTTTTTCTTACCATGAAAAGAAAT 2040
Db 1981 AACAAATCAGGACCATTCGGTTCGTATAGTGACTGCTTTTTTCTTACCATGAAAAGAAAT 2040
Qy 2041 CAGATTTTGCATATGACTTTAGCAACCTCAGCAGTGGGCTCATTAATGATGGCCCC 2100
Db 2041 CAGATTTTGCATATGACTTTAGCAACCTCAGCAGTGGGCTCATTAATGATGGCCCC 2100
Qy 2101 AGCTTCCACTCCAAAGGAACAAATACTTCCATTTCTTCAATATCAGTTTATGTTGGCAT 2160
Db 2101 AGCTTCCACTCCAAAGGAACAAATACTTCCATTTCTTCAATATCAGTTTATGTTGGCAT 2160
Qy 2161 GAGGGGAAGAAGATGCTCTCTGTACCAACAAATATACAGACTTTTACAGTAAAAAGAAATA 2220
Db 2161 GAGGGGAAGAAGATGCTCTCTGTACCAACAAATATACAGACTTTTACAGTAAAAAGAAATA 2220

```
QY 2221 GTGGAGGTCAGATGATTACAAATTTGGTAGGGGCAATTTGTATGCCAGTCAACAATT 2280
DB 2221 GTGGAGGTCAGATGATTACAAATTTGGTAGGGGCAATTTGTATGCCAGTCAACAATT 2280
QY 2281 ATTCTCTCTGAAAGTAAGGTTTCCGAGAGCCCTTATCATCAAAATCCATCATCTTGCGCA 2340
DB 2281 ATTCTCTCTGAAAGTAAGGTTTCCGAGAGCCCTTATCATCAAAATCCATCATCTTGCGCA 2340
QY 2341 GATACATTCATAGGAGTCACAGTTGAAACCCACATTTGAAAAATATTAAATATAAAAGAAAT 2400
DB 2341 GATACATTCATAGGAGTCACAGTTGAAACCCACATTTGAAAAATATTAAATATAAAAGAAAT 2400
QY 2401 ATGTTCCCAAGTTCCAAACAGCCAAATACAGATGTGCATTTCTTTTATAAGTCTTCTACA 2460
DB 2401 ATGTTCCCAAGTTCCAAACAGCCAAATACAGATGTGCATTTCTTTTATAAGTCTTCTACA 2460
QY 2461 GCAACAACATCTTGTATTAAATGGCCGATCACTGCTGTGAAAAATGAGGTGTAATCCTACT 2520
DB 2461 GCAACAACATCTTGTATTAAATGGCCGATCACTGCTGTGAAAAATGAGGTGTAATCCTACT 2520
QY 2521 AAATCTGGAGCAGGAGTGTATTTCACTCCCCACCAAGTGCCAGCAGGTACCTGTGATGGG 2580
DB 2521 AAATCTGGAGCAGGAGTGTATTTCACTCCCCACCAAGTGCCAGCAGGTACCTGTGATGGG 2580
QY 2581 TGTACGTTCTATTTCTGTGGGAGAGTGTGAAGCTTGCCTCTGTGTACGGAGCATGAC 2640
DB 2581 TGTACGTTCTATTTCTGTGGGAGAGTGTGAAGCTTGCCTCTGTGTACGGAGCATGAC 2640
QY 2641 TTCCATGAGATTGAGGAGCCCTGCAAGAGAGGATTTTCAGGAAACCTTGTATGTGGAAAT 2700
DB 2641 TTCCATGAGATTGAGGAGCCCTGCAAGAGAGGATTTTCAGGAAACCTTGTATGTGGAAAT 2700
QY 2701 GAACCTAAATGTGCATTAAGGAATTTCTTTCCTGAGAAAAAGTTGCAACCTGTGAA 2760
DB 2701 GAACCTAAATGTGCATTAAGGAATTTCTTTCCTGAGAAAAAGTTGCAACCTGTGAA 2760
QY 2761 ACGGTTGACTTTTGGTGAAGTGGGAGCGGTGTGGAGCTTTTACTCCCGTTTGTGCTG 2820
DB 2761 ACGGTTGACTTTTGGTGAAGTGGGAGCGGTGTGGAGCTTTTACTCCCGTTTGTGCTG 2820
QY 2821 GTGGCTCTGACCTGTACTTCTTGAAAAAGAAATCAAAAATGGAATATATCCAAAG 2880
DB 2821 GTGGCTCTGACCTGTACTTCTTGAAAAAGAAATCAAAAATGGAATATATCCAAAG 2880
QY 2881 TTAGTAATCAGCAGTAACCTCAAGAGTGTGAATCCCGGCTGCAGACAGTTGTGCTATC 2940
DB 2881 TTAGTAATCAGCAGTAACCTCAAGAGTGTGAATCCCGGCTGCAGACAGTTGTGCTATC 2940
QY 2941 ATGGAAGGAGAAGATAATGAAGAGGAAGTTGTATATTTCCAATAAACAGTCACTACTAGGA 3000
DB 2941 ATGGAAGGAGAAGATAATGAAGAGGAAGTTGTATATTTCCAATAAACAGTCACTACTAGGA 3000
QY 3001 AAACCTCAAAATCTTTGGCAACCAAGGAAAAAGAACCAATTTTGAATCTCTTCAACTGAAA 3060
DB 3001 AAACCTCAAAATCTTTGGCAACCAAGGAAAAAGAACCAATTTTGAATCTCTTCAACTGAAA 3060
QY 3061 ACCTCAAGATCCCAATATATCAAGAGACAGTGTGTAGCCCTTGAGACTAATGAACAAA 3120
DB 3061 ACCTCAAGATCCCAATATATCAAGAGACAGTGTGTAGCCCTTGAGACTAATGAACAAA 3120
QY 3121 GAAACCTGCTCTAGTTTACAGGACCATATTTTAGGCTGTCTCTCATACCTGTGCATTT 3180
DB 3121 GAAACCTGCTCTAGTTTACAGGACCATATTTTAGGCTGTCTCTCATACCTGTGCATTT 3180
QY 3181 GGTGATCTCACAGAGAGGGCCATGCCGTGAAAAGGGAAGGAGATTGAAACATTTTGATT 3240
DB 3181 GGTGATCTCACAGAGAGGGCCATGCCGTGAAAAGGGAAGGAGATTGAAACATTTTGATT 3240
QY 3241 GCCTTATCATGTTCAAGTACCTTGCCCAAAATAAGGAAAGCAATGATTTGGTCTCAA 3300
DB 3241 GCCTTATCATGTTCAAGTACCTTGCCCAAAATAAGGAAAGCAATGATTTGGTCTCAA 3300
```

```
QY 3301 CTGAAGATGAAGCTCAACTCAGGAAGAGATTATCTGTATATACACATAACTGAAAAACA 3360
DB 3301 CTGAAGATGAAGCTCAACTCAGGAAGAGATTATCTGTATATACACATAACTGAAAAACA 3360
QY 3361 AGTTTAAAGCCCAACCAATGCACCTGCTGATGCATGCCATATAAATTAATGGTAACCTTTATT 3420
DB 3361 AGTTTAAAGCCCAACCAATGCACCTGCTGATGCATGCCATATAAATTAATGGTAACCTTTATT 3420
QY 3421 CTTTATGATGCTACATACACAAAGTGTGATTTGGAAGGCACATGTGAGCATATGCATTATG 3480
DB 3421 CTTTATGATGCTACATACACAAAGTGTGATTTGGAAGGCACATGTGAGCATATGCATTATG 3480
QY 3481 ATCCAAATTTATCTTTTCTTTTGTATATTTTGGGAAAAATTTAAATTTTTTAAAGGTA 3540
DB 3481 ATCCAAATTTATCTTTTCTTTTGTATATTTTGGGAAAAATTTAAATTTTTTAAAGGTA 3540
QY 3541 AAAAAAATAAAAAA 3556
DB 3541 AAAAAAATAAAAAA 3556

RESULT 2
US-10-140-164-1
; Sequence 1, Application US/10140164
; Publication No. US20030072736A1
; GENERAL INFORMATION:
; APPLICANT: Baker et al.
; TITLE OF INVENTION: Human Tumor Necrosis Factor Receptor TR16
; FILE REFERENCE: PFS14C1
; CURRENT APPLICATION NUMBER: US/10/140,164
; CURRENT FILING DATE: 2002-05-08
; PRIOR APPLICATION NUMBER: 09/637,856
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/148,348
; PRIOR FILING DATE: 1999-08-12
; PRIOR APPLICATION NUMBER: 60/148,683
; PRIOR FILING DATE: 1999-08-13
; PRIOR APPLICATION NUMBER: 60/148,870
; PRIOR FILING DATE: 1999-08-13
; PRIOR APPLICATION NUMBER: 60/148,758
; PRIOR FILING DATE: 1999-08-16
; PRIOR APPLICATION NUMBER: 60/149,181
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: 60/149,453
; PRIOR FILING DATE: 1999-08-18
; PRIOR APPLICATION NUMBER: 60/149,498
; PRIOR FILING DATE: 1999-08-19
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 3390
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2892)
US-10-140-164-1

Query Match 90.4%; Score 3214; DB 9; Length 3390;
Best Local Similarity 95.3%; Pred. No. 0;
Matches 3390; Conservative 0; Mismatches 0; Indels 166; Gaps 1;
```

```
QY 1 ATGCTGTTCCGCGCCGCGGCGGTACGGGCGCAGGGGCTGGGGGCGCGCGGAGGCT 60
DB 1 ATGCTGTTCCGCGCCGCGGCGGCGGTACGGGCGCAGGGGCTGGGGGCGCGCGGAGGCT 60
QY 61 CCCGCGCGCGGCGCTGCGCGCCCTGGAGCCCGCTGCTGCTGGGCGCTCGCC 120
DB 61 CCCGCGCGCGGCGCTGCGCGCCCTGGAGCCCGCTGCTGCTGGGCGCTCGCC 120
QY 121 GGCTGCAGCGGCGCTGGGCTGGGGACCTGCCCTCCTCCAGCCGCGCTTCCTCT 180
DB 121 GGCTGCAGCGGCGCTGGGCTGGGGACCTGCCCTCCTCCAGCCGCGCTTCCTCTCT 180
```

QY 181 TCCAGGAGAAAGATTATCACTTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGG 240
DB 181 TCCAGGAGAAAGATTATCACTTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGG 240
QY 241 AGAGTTGGCAATCCAAATTCGACGTGGACGTGCTCTGGCTGCCCTGACCCAGTGAGAGCC 300
DB 241 AGAGTTGGCAATCCAAATTCGACGTGGACGTGCTCTGGCTGCCCTGACCCAGTGAGAGCC 300
QY 301 AAAGATGACATTTCTCCCTGCTCTCTGGAGAGTATCTAGAAATGAAGAACACAGGTATGC 360
DB 301 AAAGATGACATTTCTCCCTGCTCTCTGGAGAGTATCTAGAAATGAAGAACACAGGTATGC 360
QY 361 AGTAAGTGTGGTGAAGGACCACTATTCTTGGCCAGTGGCATCAAAATTTGATGAATGGGAT 420
DB 361 AGTAAGTGTGGTGAAGGACCACTATTCTTGGCCAGTGGCATCAAAATTTGATGAATGGGAT 420
QY 421 GAATTCGGCCGAGGATTTCTAACATCGCAACATTCATGGGACACTGTGGTGGCCCTTCT 480
DB 421 GAATTCGGCCGAGGATTTCTAACATCGCAACATTCATGGGACACTGTGGTGGCCCTTCT 480
QY 481 GACAGCAGCCAGAGCGGTGTAAACACTCTTCTTGGATCCCTCGTGGAACTTACATAGAA 540
DB 481 GACAGCAGCCAGAGCGGTGTAAACACTCTTCTTGGATCCCTCGTGGAACTTACATAGAA 540
QY 541 TCTAATCGTGATGACGTGACCGGTGTCTTGTATGCTATGCTGACCTTTAAGAAGTCAGGC 600
DB 541 TCTAATCGTGATGACGTGACCGGTGTCTTGTATGCTATGCTGACCTTTAAGAAGTCAGGC 600
QY 601 TATGCTCTCTTGTAGTACAGTATCGGACACACATCTCTTTCAGTTCCTTTATTCAA 660
DB 601 TATGCTCTCTTGTAGTACAGTATCGGACACACATCTCTTTCAGTTCCTTTATTCAA 660
QY 661 AATGATCAGTGCAGGAGATGACACACCACTGACAGTGGGTAAAATTTACAGACAA 720
DB 661 AATGATCAGTGCAGGAGATGACACACCACTGACAGTGGGTAAAATTTACAGACAA 720
QY 721 GGAGAAATGGGGTCTCATTTCTGTAATGCTGAAATCAGGCACAAACATCTCTACTGGAGA 780
DB 721 GGAGAAATGGGGTCTCATTTCTGTAATGCTGAAATCAGGCACAAACATCTCTACTGGAGA 780
QY 781 ACTACAGGCATCCTTATGGGTTCTAAGGGGTTCAAGCCCTGCTGTAATAAATATCACA 840
DB 781 ACTACAGGCATCCTTATGGGTTCTAAGGGGTTCAAGCCCTGCTGTAATAAATATCACA 840
QY 841 ATTGAAGGGTGGCGTACACATCAGAAATGTTTTCCTTCCAAGCCAGGCACATTCAGCAAC 900
DB 841 ATTGAAGGGTGGCGTACACATCAGAAATGTTTTCCTTCCAAGCCAGGCACATTCAGCAAC 900
QY 901 AAACCAAGTTTCATTCNACTGCCAGGTGTGTCGCCAGAAACACCTATTCTCAGAAAGGACC 960
DB 901 AAACCAAGTTTCATTCNACTGCCAGGTGTGTCGCCAGAAACACCTATTCTCAGAAAGGACC 960
QY 961 AAAGAAATGTAAGGTGTAAGACGACTCTCAATTTTCAGGATCCAGTCAAGTGTACAGAG 1020
DB 961 AAAGAAATGTAAGGTGTAAGACGACTCTCAATTTTCAGGATCCAGTCAAGTGTACAGAG 1020
QY 1021 CGCCCTCCCTGTACCAAAAGACTATTTCAGATCCATCTCCATGTGATGAAGAAGGA 1080
DB 1021 CGCCCTCCCTGTACCAAAAGACTATTTCAGATCCATCTCCATGTGATGAAGAAGGA 1080
QY 1081 AAGACACAGATATGTACAAAGTGGATAGAGCCCAAAATCTGCCGGAGGATCTCACAGAT 1140
DB 1081 AAGACACAGATATGTACAAAGTGGATAGAGCCCAAAATCTGCCGGAGGATCTCACAGAT 1140
QY 1141 GCTATTAGATTCGCCCTTCTCGAGAGAAGAGGATTGTCCGCTTGGCAACCTCGGATTT 1200
DB 1141 GCTATTAGATTCGCCCTTCTCGAGAGAAGAGGATTGTCCGCTTGGCAACCTCGGATTT 1200
QY 1201 TATAACAATGGATCATCTTCTTGGCCATCCCTGTCCCTCGGAAACATTTTCAGATGGAACC 1260
DB 1201 TATAACAATGGATCATCTTCTTGGCCATCCCTGTCCCTCGGAAACATTTTCAGATGGAACC 1260

QY 1261 AAAGAAATGTAGACCATGTCCAGCAGAAACGGAGCCTGCACCTTGGCTTTGAATATAAATGG 1320
DB 1261 AAAGAAATGTAGACCATGTCCAGCAGAAACGGAGCCTGCACCTTGGCTTTGAATATAAATGG 1320
QY 1321 TCGAAATGCTCTTCCCTGGCAACATGAACACTTCCCTCAATGTTGGGAATTCAAAGTCC 1380
DB 1321 TCGAAATGCTCTTCCCTGGCAACATGAACACTTCCCTCAATGTTGGGAATTCAAAGTCC 1380
QY 1381 GATGGAATGAATGGTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGCTGGAGTTCCT 1440
DB 1381 GATGGAATGAATGGTGGGAGGTGGCTGGAGATCATATCCAGAGTGGGCTGGAGTTCCT 1440
QY 1441 GACAAATGATTACTGATCTTAACTTGCAGATATCCAGAGTAAACCCACCACTATG 1500
DB 1441 GACAAATGATTACTGATCTTAACTTGCAGATATCCAGAGTAAACCCACCACTATG 1500
QY 1501 ACTGAGGCCACGGTTCGAACTAGGAAGAAATACAAATTTGCTCTTTCAGACCCCTCTGTTCA 1560
DB 1501 ACTGAGGCCACGGTTCGAACTAGGAAGAAATACAAATTTGCTCTTTCAGACCCCTCTGTTCA 1560
QY 1561 GCTGACTGTGTTTGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1620
DB 1561 GCTGACTGTGTTTGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1620
QY 1621 TCGTGGGTGGAAACCAAGAAACAAAGCTTACACCCATATCATCTTCAAGAAATGCAACT 1680
DB 1621 TCGTGGGTGGAAACCAAGAAACAAAGCTTACACCCATATCATCTTCAAGAAATGCAACT 1680
QY 1681 TTTTACATTTTACATGGGCTTCCAGAGAACTAATCAGGGTCAAGATAATACACGGTTCACTC 1740
DB 1681 TTTTACATTTTACATGGGCTTCCAGAGAACTAATCAGGGTCAAGATAATACACGGTTCACTC 1740
QY 1741 AATGACATGGTGAAGATTTTATTTCTATCAGCCACTTAATGCAAGTTGATGGGTGGCGTCC 1800
DB 1741 AATGACATGGTGAAGATTTTATTTCTATCAGCCACTTAATGCAAGTTGATGGGTGGCGTCC 1800
QY 1801 TCATGCCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860
DB 1801 TCATGCCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860
QY 1861 CCAGGCCACTTACATTTGAGAAAGAAACCAACCACTGCAAGGAATGTCCACCTGACACCTAC 1920
DB 1861 CCAGGCCACTTACATTTGAGAAAGAAACCAACCACTGCAAGGAATGTCCACCTGACACCTAC 1920
QY 1921 CTGTCCATACATCAGTCTATGGCAAGAGGCTTGTATTTCCATGCGGGCTGGGAGTAA 1980
DB 1921 CTGTCCATACATCAGTCTATGGCAAGAGGCTTGTATTTCCATGCGGGCTGGGAGTAA 1980
QY 1981 AACAAATCAGGACCTTCCGGTTTCTGCTATAGTACTGCTTTTCTTCTTCTTCTTCTTCTTCTTCT 2040
DB 1981 AACAAATCAGGACCTTCCGGTTTCTGCTATAGTACTGCTTTTCTTCTTCTTCTTCTTCTTCTTCT 2040
QY 2041 CAGATTTTCCACTATGACTTTTACCAACCTCAGCAGTGTGGGCTCATTAATGAATGGCCCT 2100
DB 2041 CAGATTTTCCACTATGACTTTTACCAACCTCAGCAGTGTGGGCTCATTAATGAATGGCCCT 2100
QY 2101 AGCTTTCACCTCCAAAGGAACAAATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2160
DB 2101 AGCTTTCACCTCCAAAGGAACAAATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2160
QY 2161 GAGGGGAAGAAGATGGCTCTGTAACCAATATACAGACTTTTACAGTAAAGAAATA 2220
DB 2161 GAGGGGAAGAAGATGGCTCTGTAACCAATATACAGACTTTTACAGTAAAGAAATA 2220
QY 2221 GTGGCAGGCTCAGATGATTTACAAATTTTGGTAGGGCAATTTGATGCGCAGTCAACAAT 2280
DB 2221 GTGGCAGGCTCAGATGATTTACAAATTTTGGTAGGGCAATTTGATGCGCAGTCAACAAT 2280
QY 2281 ATTTCCTTCTGAAGTAAGGGTTTCCAGCAGGCTTATCATCATCAATTCATCATCTTCTTCTTCTTCT 2340
DB 2281 ATTTCCTTCTGAAGTAAGGGTTTCCAGCAGGCTTATCATCATCAATTCATCATCTTCTTCTTCTTCT 2340
QY 2341 GATACATTCATAGGAGTCACTGAGTTGAAACCACTTCAAAATATTAATATATAAAGAAAT 2400

|||||
Db 2341 GATACATTATAGGAGTACAGTTGAACACCATTTGAATAATTAATAAAGAAAT 2400
QY 2401 ATGTTCCAGTTCCCAAGCCAAATACAGATGTGCATTTCTTTTATAAGTCTTCTACA 2460
Db 2401 ATGTTCCAGTTCCCAAGCCAAATACAGATGTGCATTTCTTTTATAAGTCTTCTACA 2460
QY 2461 GCACCAACATCTGTATTAATGCGCGATCACTGCTGCAAAATGAGGTGTAATCCTACT 2520
Db 2461 GCACCAACATCTGTATTAATGCGCGATCACTGCTGCAAAATGAGGTGTAATCCTACT 2520
QY 2521 AAATCTGGAGAGGAGTGTATTCAGTCCCGACCAAGTGCACAGGACCTGTGTGATGGG 2580
Db 2521 AAATCTGGAGAGGAGTGTATTCAGTCCCGACCAAGTGCACAGGACCTGTGTGATGGG 2580
QY 2581 TGTACGTTCTATTTCTGTGGAGAGTGTGAGCTTGCCCTCTGTGTACGAGGATGAC 2640
Db 2581 TGTACGTTCTATTTCTGTGGAGAGTGTGAGCTTGCCCTCTGTGTACGAGGATGAC 2640
QY 2641 TTCCATGAGATTGAGGAGCCTGCAAGAGAGATTTCAGAAACCTTGTATGTGTGGAAT 2700
Db 2641 TTCCATGAGATTGAGGAGCCTGCAAGAGAGATTTCAGAAACCTTGTATGTGTGGAAT 2700
QY 2701 GAACCTAAATGTGTCAATTAAGGAATTTCTTTGCCCTGAGAAAAGTTGCCAACCTGTGAA 2760
Db 2701 GAACCTAAATGTGTCAATTAAGGAATTTCTTTGCCCTGAGAAAAGTTGCCAACCTGTGAA 2760
QY 2761 ACGGTTGACTTTTGGCTGAAGTGGAGCGCGTGGGAGCTTTTACTGCCGTTTGGTG 2820
Db 2761 ACGGTTGACTTTTGGCTGAAGTGGAGCGCGTGGGAGCTTTTACTGCCGTTTGGTG 2820
QY 2821 GTGGCTGTACCTGTCTACTTCTGCAAAAAGAAATCAAAAAGTGAATACAAATATTTCAAAG 2880
Db 2821 GTGGCTGTACCTGTCTACTTCTGCAAAAAGAAATCAAAAAGTGAATACAAATATTTCAAAG 2880
QY 2881 TTAGTAATGACGACTAACTCAAAAGAGTGTGAACCTCCGCGTGCAGACAGTTGTGTATC 2940
Db 2859 ----- 2858
QY 2941 ATGGAAGGAGAAGATATGAAGAGAGTGTGTATATTTCCAATAACACGACTACTAGGA 3000
Db 2859 ----- 2858
QY 3001 AAACCTAAATCTTTGGCAACCAAGGAAAAGAGACATTTTGAATCTGTTCACCTGAAA 3060
Db 2859 -----GAAAAAGAGACATTTTGAATCTGTTCACCTGAAA 2894
QY 3061 ACCTCAAGATCCCAATATATGAAGAGACAGTGTGTAGCCCTTGAGACTAATGAACAAA 3120
Db 2895 ACCTCAAGATCCCAATATATGAAGAGACAGTGTGTAGCCCTTGAGACTAATGAACAAA 2954
QY 3121 GAAACCTGCTCTAGTTTTCAGGACCATATTTTAGGCTGTCTCTCATACCTGTCACT 3180
Db 2955 GAAACCTGCTCTAGTTTTCAGGACCATATTTTAGGCTGTCTCTCATACCTGTCACT 3014
QY 3181 GGTGATCTCAGAGGAGGCCATGCGCGTGAAAAGGAAGAGAGATTGAAACATTTGATT 3240
Db 3015 GGTGATCTCAGAGGAGGCCATGCGCGTGAAAAGGAAGAGAGATTGAAACATTTGATT 3074
QY 3241 GCCTTATACATGCTCAAGTACCTTGCCAAATGAAGGAAGCAATGATTGGGTCTCAA 3300
Db 3075 GCCTTATACATGCTCAAGTACCTTGCCAAATGAAGGAAGCAATGATTGGGTCTCAA 3134
QY 3301 CTGAAGATGAAGCTCAACTCAGGAAGAGATTATCTGTATATACATAACTGAAACCA 3360
Db 3135 CTGAAGATGAAGCTCAACTCAGGAAGAGATTATCTGTATATACATAACTGAAACCA 3194
QY 3361 AGTTTAAGCCCAACCAATGCACTGCTGTATGATGCCATATTAATTAATGGGTAACTTTTATT 3420
Db 3195 AGTTTAAGCCCAACCAATGCACTGCTGTATGATGCCATATTAATTAATGGGTAACTTTTATT 3254
QY 3421 CTTTATGATGCTACATAACAAAGTGTGATTTGGAGGACCATGTGAGCATATGTCATTATG 3480
|||||



Db 3255 CTTTATGATGTCTACATAACAAGTGTGATTTGGAAGGCACATGTGACATATGCAATTATG 3314
QY 3481 ATCCAATTTATGTTTCTTTTCTTTTATTTTCTGGGAAAATTTAAATTTTTTTTAAAGTA 3540
Db 3315 ATCCAATTTATGTTTCTTTTCTTTTATTTTCTGGGAAAATTTAAATTTTTTTTAAAGTA 3374
QY 3541 AAAAAAATAAAAAA 3556
Db 3375 AAAAAAATAAAAAA 3390

RESULT 3
US-10-002-050-19
; Sequence 19, Application US/10002050
; Publication No. US20030032095A1
; GENERAL INFORMATION:
; APPLICANT: Shinketsu, Richard
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corine
; APPLICANT: Yang, Meljia
; APPLICANT: Boldog, Ferenc
; APPLICANT: Herrmann, John
; TITLE OF INVENTION: No. US20030032095A1el Nucleic Acid Sequences Encoding Human Se
; FILE REFERENCE: 15966-554 Cura-54 CON-514
; CURRENT APPLICATION NUMBER: US/10/002.050
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 09/604, 286
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/140, 584
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 1737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (296)..(1687)
; NAME/KEY: variation
; LOCATION: (1)..(1737)
; OTHER INFORMATION: N may be any nucleotide
US-10-002-050-19

Query Match 42.2%; Score 1499.6; DB 9; Length 1737;
Best Local Similarity 98.7%; Pred. No. 0;
Matches 1545; Conservative 0; Mismatches 10; Indels 10; Gaps 3;

QY 189 GAAAGATTATCACTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGAGAGTTGC 248
Db 142 GAAAGATTATCACTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGAGAGTTGC 201
QY 249 CATTCCAAATTCGACGTGACCTGCTGCGCTGACCCAGTGAGAGGCAAGAATG 308
Db 202 CATTCCAAATTCGACGTGACCTGCTGCGCTGACCCAGTGAGAGGCAAGAATG 261
QY 309 CACTTTCTCTGCTCTCTGGAGAGTATCTAGAAATAGAAACCCAGGTATGCAAGTAAGTG 368
Db 262 CACTTTCTCTGCTCTCTGGAGAGTATCTAGAAATAGAAACCCAGGTATGCAAGTAAGTG 321
QY 369 TGGTGAAGGACCATATTCCTTGGGAGTGGCATCAAAATTTGATGAATGGATGAATGGC 428
Db 322 TGGTGAAGGACCATATTCCTTGGGAGTGGCATCAAAATTTGATGAATGGATGAATGGC 381
QY 429 GGAGGATTTTCTTAACATCGCAACATTCATGGACACTGTGGTGGGCCCTTCTGACAGCAG 488
Db 382 GGAGGATTTTCTTAACATCGCAACATTCATGGACACTGTGGTGGGCCCTTCTGACAGCAG 441
QY 489 GCAGAGCGGTGTAAACACTCTTTCTTTGGATCCCTCTGGAACACTACATAGAATCTAATCG 548
Db 442 GCAGAGCGGTGTAAACACTCTTTCTTTGGATCCCTCTGGAACACTACATAGAATCTAATCG 501
QY 549 TGATGACTGCACGGTGTCTTTTGATCTATGCTGTGCACCTTAAAGAGTCAGGCTATGCTTT 608
|||||

Db 502 TGATGACGTCACGGGTCTTTGATCTATGCTGTGCACCTTAAGAACTCAGGCTATGCTTT 561
QY 609 CTTTGAGTACCACTATGTCGACAAACATCTTCTTTGAGTCTCTTTATTCAAAATGATCA 668
Db 562 CTTTGAGTACCACTATGTCGACAAACATCTTCTTTGAGTCTCTTTATTCAAAATGATCA 621
QY 669 GTGCCAGGAGATGGACACCACCTGACAAAGTGGTAAACTTTACAGACAATGGAGAATG 728
Db 622 GTGCCAGGAGATGGACACCACCTGACAAAGTGGTAAACTTTACAGACAATGGAGAATG 681
QY 729 GGGCTCTCATTCTGTAATGCTGAATCAGGCACAAACATCTACTCTGAGAACTACAGG 788
Db 682 GGGCTCTCATTCTGTAATGCTGAATCAGGCACAAACATCTACTCTGAGAACTACAGG 741
QY 789 CATCCCTTATGGGTCTTAAGCGGTCAAGCGTGTGCTGGTAAATAATACAAATTCGAAG 848.
Db 742 CATCCCTTATGGGTCTTAAGCGGTCAAGCGTGTGCTGGTAAATAATACAAATTCGAAG 801
QY 849 GGTGGGTACACATCAGATGTTTTCTTGCAGCCAGGCACATTCAGCAACAACCCAGG 908
Db 802 GGTGGGTACACATCAGATGTTTTCTTGCAGCCAGGCACATTCAGCAACAACCCAGG 861
QY 909 TTCATTCAACTGCCAGGTGTCCCAAGAACCTTATCTGAGAAAGGAGCCAAAGAAATG 968
Db 862 TTCATTCAACTGCCAGGTGTCCCAAGAACCTTATCTGAGAAAGGAGCCAAAGAAATG 921
QY 969 TATAAGGTGAAGAGACCTCTCAATTTTC - - - - - AGGATCCAGTGAGTACAGAGCG 1022
Db 922 TATAAGGTGAAGAGACCTCTCAATTTTCAGAGGAAGGATCCAGTGAGTACAGAGCG 981
QY 1023 CCCTCCCTGTACCACAAAGACATTTTCCAGATCCATCTCCATGATGATGAAGAAGAA 1082
Db 982 CCCTCCCTGTACCACAAAGACATTTTCCAGATCCATCTCCATGATGATGAAGAAGAA 1041
QY 1083 GACACAGATAATACAAAGTGGATAGAGCCCAAAATCTCCGGGAGGATCTCACAGATGC 1142
Db 1042 GACACAGATAATACAAAGTGGATAGAGCCCAAAATCTCCGGGAGGATCTCACAGATGC 1101
QY 1143 TATTAGATGCCCTTCTGGAGAGAAAGGATTTCCGCCCTTGCAACCCCTGGATTTTA 1202
Db 1102 TATTAGATGCCCTTCTGGAGAGAAAGGATTTCCGCCCTTGCAACCCCTGGATTTTA 1161
QY 1203 TAACAATGGATCATCTTTGTCATCCCTGCTCCTCGTGAACATTTTCAGATGGAACCAA 1262
Db 1162 TACATGGATCATCTTTGTCATCCCTGCTCCTCGTGAACATTTTCAGATGGAACCAA 1221
QY 1263 AGAATGTAGACCATGTCCAGCAGGAACGGAGCCTGACATTTGGCTTTGAATATAAATGGTG 1322
Db 1222 AGAATGTAGACCATGTCCAGCAGGAACGGAGCCTGACATTTGGCTTTGAATATAAATGGTG 1281
QY 1323 GAATGTCTTCCTGGCAACATGAAACTTTCCTGCTCAATGTTGGGAATTCAAAGTGCGA 1382
Db 1282 GAATGTCTTCCTGGCAACATGAAACTTTCCTGCTCAATGTTGGGAATTCAAAGTGCGA 1341
QY 1383 TGGGAATGAATGTTGGGAGGTGGCTGGAGATCATATCCAGATGGGGCTGGAGGTTCTCA 1442
Db 1342 TGGGAATGAATGTTGGGAGGTGGCTGGAGATCATATCCAGATGGGGCTGGAGGTTCTCA 1401
QY 1443 CAATGATACCTGATCTTAACTTGCATATCCAGGATTTAAACCAACCAACATCTATGAC 1502
Db 1402 CAATGATACCTGATCTTAACTTGCATATCCAGGATTTAAACCAACCAACATCTATGAC 1461
QY 1503 TGGAGCCAGGGTCTGAACTAGGAGATTAACATTTGCTTTGAGCCCTCTGTTCCAGC 1562
Db 1462 TGGAGCCAGGGTCTGAACTAGGAGATTAACATTTGCTTTGAGCCCTCTGTTCCAGC 1521
QY 1563 TGACTGTGTTTTGACTTCCATGGTGGATTAATAGAAAAAGTACAAATGTGTTAGAATC 1622
Db 1522 TGACTGTGTTTTGACTTCCATGGTGGATTAATAGAAAAAGTACAAATGTGTTAGAATC 1581
QY 1623 GTGGGTGCAACCAAGAAACAGCTTTACACCCATATCATCTTCAAGAAATGCAACTTT 1682
|||||

Db 1582 GTGGGTGCAACCAAGAAAAAAGCTTTACACCCATATCATCTTCAAGAAATGCAACTTT 1641
QY 1683 TACATTTACATGGG--CATTTCCAGAGAACTAAT--CAGGGTCAAGNATATAGACGGTTCA 1738
Db 1642 TACATTTACATGGGCAATTTCCAGAGAACTAATTCAGGGTCCAAAGATATAGACGGTTCC 1701
QY 1739 TCAAT 1743
Db 1702 NCCAT 1706
|||
RESULT 4
US-10-002-304-19
; Sequence 19, Application US/10002304
; Publication No. US20030036185A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernnet, Corine
; APPLICANT: Yang, Meljia
; APPLICANT: Boldog, Ferenc
; APPLICANT: Herrmann, John
; TITLE OF INVENTION: Polynucleotides and polypeptides encoded thereby
; FILE REFERENCE: 15866-534 Cura-54 CON-58
; CURRENT APPLICATION NUMBER: US/10/002.304
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 09/604,286
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/140,584
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 1737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (296)..(1687)
; NAME/KEY: variation
; LOCATION: (1)..(1737)
; OTHER INFORMATION: N may be any nucleotide
US-10-002-304-19
Query Match 42.2%; Score 1499.6; DB 9; Length 1737;
Best Local Similarity 98.7%; Pred. No. 0;
Matches 1545; Conservative 0; Mismatches 10; Indels 10; Gaps 3;

QY 189 GAAAGATTATCACTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGGAGAGTTGC 248
Db 142 GAAAGATTATCACTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGGAGAGTTGC 201
QY 249 CATTCCAAATTCGAGTGGAGTGTCTGGGCTGCTGACCCAGTGGAGAGGCAAGAATG 308
Db 202 CATTCCAAATTCGAGTGGAGTGTCTGGGCTGCTGACCCAGTGGAGAGGCAAGAATG 261
QY 309 CACTTTCTCTGCTGCTTCTGGAGAGTATCTAGAAATGAAGAACCCAGGTATGCAGTAAGTG 368
Db 262 CACTTTCTCTGCTGCTTCTGGAGAGTATCTAGAAATGAAGAACCCAGGTATGCAGTAAGTG 321
QY 369 TGGTGAAGGCACCTATTCCTTTGGGAGTGGCATCAATTTGATGAATGGGATGAATTTGCC 428
Db 322 TGGTGAAGGCACCTATTCCTTTGGGAGTGGCATCAATTTGATGAATGGGATGAATTTGCC 381
QY 429 GGCAGGATTTTCTAACATCGCAACATTCATGGACACTGTGGTGGGCCCTTCTGCACAGCAG 488
Db 382 GGCAGGATTTTCTAACATCGCAACATTCATGGACACTGTGGTGGGCCCTTCTGCACAGCAG 441
QY 489 GCCAGACGGCTGTAACTACTCTTTTGGATCCCTGCTGGAACTACATAGAAATCTAATTCG 548
Db 442 GCCAGACGGCTGTAACTACTCTTTTGGATCCCTGCTGGAACTACATAGAAATCTAATTCG 501
QY 549 TGATGACTGCACGGTGTCTTTGATCTATGCTGTGCACCTTAAAGAAAGTCAGGCTATGCTTT 608
|||||

```
|||||
Db 502 TGATGACTGACGGTCTCTTTGATCTATGCTGTGCACCTTAAAGAGTCAGGCTATGCTCTT 561
QY 609 CTTTGAGTACCAGTATGTCGACAACAACATCTCTTTGAGTCTTTATTCAAAATGATCA 668
Db 562 CTTTGAGTACCAGTATGTCGACAACAACATCTCTTTGAGTCTTTATTCAAAATGATCA 621
QY 669 GTGCCAGGAGATGGACACCAACCACTGACAAAGTGGGTAAAACTTACAGACAATGGAGAATG 728
Db 622 GTGCCAGGAGATGGACACCAACCACTGACAAAGTGGGTAAAACTTACAGACAATGGAGAATG 681
QY 729 GGGCTCTCAATCTGTAATGCTGAATTCAGGCACAACATACCTACTCTGAGAACTACAGG 788
Db 682 GGGCTCTCAATCTGTAATGCTGAATTCAGGCACAACATACCTACTCTGAGAACTACAGG 741
QY 789 CATCTCTTATGGCTTCTTAAGCGGTCAAGCCTGTGCTGTTAAAAATATACAAATTTGAAG 848
Db 742 CATCTCTTATGGCTTCTTAAGCGGTCAAGCCTGTGCTGTTAAAAATATACAAATTTGAAG 801
QY 849 GGTGGCGTACACATCAGAAATGTTTTCCTTTCGAAGCCAGGCACATTCAGCAACAACCCAGG 908
Db 802 GGTGGCGTACACATCAGAAATGTTTTCCTTTCGAAGCCAGGCACATTCAGCAACAACCCAGG 861
QY 909 TTCATTCAACTGCCAGGTGTGCCAGAAACACCTATTCTGAGAAAGGAGGCCAAAGAATG 968
Db 862 TTCATTCAACTGCCAGGTGTGCCAGAAACACCTATTCTGAGAAAGGAGGCCAAAGAATG 921
QY 969 TATAAGGTGTAAGACGACTCTCAATTTTC -----AGGATCCAGTGAGTACAGAGCG 1022
Db 922 TATAAGGTGTAAGACGACTCTCAATTTTCAGAGGAAGGATCCAGTGAGTACAGAGCG 981
QY 1023 CCCTCCCTGTACCACAAGAGACTATTTCAGATCCATCTCCATGTGTGATGAAGAAGGAA 1082
Db 982 CCCTCCCTGTACCACAAGAGACTATTTCAGATCCATCTCCATGTGTGATGAAGAAGGAA 1041
QY 1083 GACACAGATAATGTACAAGTGGATAGAGCCCAAAATCTGCCGGGAGGATCTCACAGATGC 1142
Db 1042 GACACAGATAATGTACAAGTGGATAGAGCCCAAAATCTGCCGGGAGGATCTCACAGATGC 1101
QY 1143 TATTAGATTGCCCTCTCGAGAGAGAGAGATGTCGCCCTTCGCAACCTCGGATTTTA 1202
Db 1102 TATTAGATTGCCCTCTCGAGAGAGAGAGATGTCGCCCTTCGCAACCTCGGATTTTA 1161
QY 1203 TAACATGGATCATCTCTTGCCATCCCTGTCTCTCTGGAACATTTTCAGATGGAAACAA 1262
Db 1162 TAACATGGATCATCTCTTGCCATCCCTGTCTCTCTGGAACATTTTCAGATGGAAACAA 1221
QY 1263 AGAATGTAGACCATGTCCAGCAGGAACGGAGCCTGCACCTTGGCTTTGAAATATAAATGGTG 1322
Db 1222 AGAATGTAGACCATGTCCAGCAGGAACGGAGCCTGCACCTTGGCTTTGAAATATAAATGGTG 1281
QY 1323 GAATGTCTCTCTGCAACATCAAAACTTCCCTGCTTCAATGTTGGAAATTCAAAGTGCA 1382
Db 1282 GAATGTCTCTCTGCAACATCAAAACTTCCCTGCTTCAATGTTGGAAATTCAAAGTGCA 1341
QY 1383 TGAATGTAATGTTGGAGGTGGCTGGAGATCATATCCAGATGGGGCTGGAGGTTCTCA 1442
Db 1342 TGAATGTAATGTTGGAGGTGGCTGGAGATCATATCCAGATGGGGCTGGAGGTTCTCA 1401
QY 1443 CAATGATTACCTGATCTTAACTTGCATATCCAGATTTAAACCAACCAATCATATGAC 1502
Db 1402 CAATGATTACCTGATCTTAACTTGCATATCCAGATTTAAACCAACCAATCATATGAC 1461
QY 1503 TGGAGCCAGGGTCTGACTAGGAGAAATACATTTGCTTTGAGACCTCTGTTCAGC 1562
Db 1462 TGGAGCCAGGGTCTGACTAGGAGAAATACATTTGCTTTGAGACCTCTGTTCAGC 1521
QY 1563 TGACTGTGTTTGTACTTTCATGGTGGATTAATAAGAAAAAGTACAAATGTGTAGAAATC 1622
Db 1522 TGACTGTGTTTGTACTTTCATGGTGGATTAATAAGAAAAAGTACAAATGTGTAGAAATC 1581
QY 1623 GTGGGTGGAACCAAGAAAAACAAGCTTTACACCCATATCATCTTCAAGAATGCAACTTT 1682
|||||
```

```
Db 1582 GTGGGTGGAACCAAGAAAAACAAGCTTACACCCATATCATCTTCAAGAATGCAACTTT 1641
QY 1683 TCATTTTACATGGG--CATTCAGAGAACTAAT--CAGGGTCAAGATAATAGACGGTTCA 1738
Db 1642 TCATTTTACATGGGCAATTCAGAGAACTAATTCAGGGTCCAGAGATAATAGACGGTTCC 1701
QY 1739 TCAAT 1743
Db 1702 NCCAT 1706

RESULT 5
US-10-003-152-19
; Sequence 19, Application US/10003152
; Patent No. US20020151494A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corine
; APPLICANT: Yang, Melija
; APPLICANT: Boldog, Ferenc
; APPLICANT: Herrmann, John
; TITLE OF INVENTION: No. US20020151494A1el Amino Acid Sequences for Human Semaphori
; FILE REFERENCE: 15966-554 Cura-54 CON-S12
; CURRENT APPLICATION NUMBER: US/10/003,152
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 09/604,286
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/140,584
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 1737
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (296)..(1687)
; NAME/KEY: variation
; LOCATION: (1)..(1737)
; OTHER INFORMATION: N may be any nucleotide
US-10-003-152-19

Query Match 42.2%; Score 1499.6; DB 12; Length 1737;
Best Local Similarity 98.7%; Pred. No. 0;
Matches 1545; Conservative 0; Mismatches 10; Indels 10; Gaps 3;
```

```
QY 189 GAAAGATTATCACTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGGAGAGTTGC 248
Db 142 GAAAGATTATCACTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGGAGAGTTGC 201
QY 249 CATTCCAAATTCGCAAGTGGACTGCTTGGCCCTGCCTGACCCAGTGGAGGCAAGAATG 308
Db 202 CATTCCAAATTCGCAAGTGGACTGCTTGGCCCTGCCTGACCCAGTGGAGGCAAGAATG 261
QY 309 CACTTCTCTCTGCTTCTGGAGAGTATCTAGAAATGAAGAACCCAGTATGCAAGTAAAGTG 368
Db 262 CACTTCTCTCTGCTTCTGGAGAGTATCTAGAAATGAAGAACCCAGTATGCAAGTAAAGTG 321
QY 369 TGGTGAAGCACTTATCTCTGGGCAAGTGGCATCAAAATTTGATGAATGGATGAATGGC 428
Db 322 TGGTGAAGCACTTATCTCTGGGCAAGTGGCATCAAAATTTGATGAATGGATGAATGGC 381
QY 429 GGCAGGATTTCTAACATCGCAACATTCATGGACACTGTGGGGCCCTTCTGACAGCAG 488
Db 382 GGCAGGATTTCTAACATCGCAACATTCATGGACACTGTGGGGCCCTTCTGACAGCAG 441
QY 489 GCCAGACGCTGTAAACAACCTCTTCTGGATCCCTCTGCGGAAACTACATAGAATCTAATCG 548
Db 442 GCCAGACGCTGTAAACAACCTCTTCTGGATCCCTCTGCGGAAACTACATAGAATCTAATCG 501
QY 549 TGATGACTGCACGGTGTCTTTTGTATCTATGCTGTGCACCTTAAAGAAGTCAGGCTATGCTTT 608
```

Db 502 TGTACTGTCACGGGTCTTTGATCTATGCTGTGCACCTTTAAGAAGTCAGGCTATGTCTTT 561
Qy 609 CTTTGAGTACCAAGTATGTGCAACAACAACATCTCTTTGAGTCTTTTATTCAAAATGATCA 668
Db 562 CTTTGAGTACCAAGTATGTGCAACAACAACATCTCTTTGAGTCTTTTATTCAAAATGATCA 621
Qy 669 GTGCCAGGAGATGGACACCAACCACTGACAAAGTGGGTAATAATACAGACAATGGAGAATG 728
Db 622 GTGCCAGGAGATGGACACCAACCACTGACAAAGTGGGTAATAATACAGACAATGGAGAATG 681
Qy 729 GGGCTCTAATCTGTAAGCTGAATTCAGGCAACAACATCTACTGGAGAACTACAGG 788
Db 682 GGGCTCTAATCTGTAAGCTGAATTCAGGCAACAACATCTACTGGAGAACTACAGG 741
Qy 789 CATCTCTATTTGGGTCTTAAGCGCGTCAAGCTGTGCTGGTAAAAATATACAAATTTGAAGG 848
Db 742 CATCTCTATTTGGGTCTTAAGCGCGTCAAGCTGTGCTGGTAAAAATATACAAATTTGAAGG 801
Qy 849 GTGGCGGTACACATCAGAAATGTTTTCTTTGCAAGCCAGGCACATTCAGCAACAACACAGG 908
Db 802 GGTGGCGTACACATCAGAAATGTTTTCTTTGCAAGCCAGGCACATTCAGCAACAACACAGG 861
Qy 909 TTCAATCAACTGCCAGGTGTGCCAGAAACACCTATTCTGAGAAAGGAGCCAAAGAATG 968
Db 862 TTCAATCAACTGCCAGGTGTGCCAGAAACACCTATTCTGAGAAAGGAGCCAAAGAATG 921
Qy 969 TATAAGGTGAAGACGACTCTCAATTTTC-----AGGACTCAGTGTACAGAGG 1022
Db 922 TATAAGGTGAAGACGACTCTCAATTTTCAGAGGAAGGATCCAGTGTACAGAGG 981
Qy 1023 CCCTCCCTGTACCAAAAAGACTATTTCCAGATCCATCTCCATGATGATGAAGAAGAA 1082
Db 982 CCCTCCCTGTACCAAAAAGACTATTTCCAGATCCATCTCCATGATGATGAAGAAGAA 1041
Qy 1083 GACACAGATAATGTACAAGTGTATAGAGCCCAAAATCTGCCGGAGGATCTCACAGATGC 1142
Db 1042 GACACAGATAATGTACAAGTGTATAGAGCCCAAAATCTGCCGGAGGATCTCACAGATGC 1101
Qy 1143 TATTAGATTGCCCTCTCTGGAGAGAAGGATTGTCCGCCCTTGCAACCCCTGGATTTTA 1202
Db 1102 TATTAGATTGCCCTCTCTGGAGAGAAGGATTGTCCGCCCTTGCAACCCCTGGATTTTA 1161
Qy 1203 TACAATGATCATCTCTTTGCCATCCCTGTCTCTCTGGAACATTTTCAGATGGAAACAA 1262
Db 1162 TAACATGATCATCTCTTTGCCATCCCTGTCTCTCTGGAACATTTTCAGATGGAAACAA 1221
Qy 1263 AGAATGTACACCATGTCCACAGCAGGAGCGCTGCACCTTGGCTTTGAATATAAATGGTG 1322
Db 1222 AGAATGTACACCATGTCCACAGCAGGAGCGCTGCACCTTGGCTTTGAATATAAATGGTG 1281
Qy 1323 GAATGTCTCTCTGGCAACATGAAACTTCCCTGCTTCAATGTTGGGAATTCAAAGTGGCA 1382
Db 1282 GAATGTCTCTCTGGCAACATGAAACTTCCCTGCTTCAATGTTGGGAATTCAAAGTGGCA 1341
Qy 1383 TGAATGAATGTTGGGAGGTGGCTGGAGATCATATCCAGATGGGGCTGGAGGTTCTGA 1442
Db 1342 TGAATGAATGTTGGGAGGTGGCTGGAGATCATATCCAGATGGGGCTGGAGGTTCTGA 1401
Qy 1443 CAATGATTACCTGTACTTAACTTCATATCCAGATCCAGATTTAAACCACCAACATCTATGAC 1502
Db 1402 CAATGATTACCTGTACTTAACTTCATATCCAGATCCAGATTTAAACCACCAACATCTATGAC 1461
Qy 1503 TGGAGCCAGGGTCTGAACTAGGAAGATAACATTTGTCTTTTGAGACCCCTCTGTTCCAG 1562
Db 1462 TGGAGCCAGGGTCTGAACTAGGAAGATAACATTTGTCTTTTGAGACCCCTCTGTTCCAG 1521
Qy 1563 TCACGTGTTTTGTACTTCATGGTGGATATTAATAGAAAAGTACAAATGTGGTAGAATC 1622
Db 1522 TGACTGTGTTTTGTACTTCATGGTGGATATTAATAGAAAAGTACAAATGTGGTAGAATC 1581
Qy 1623 GTGGGTGGAAACCAAGAAACAAGCTTACACCATATCATCTTCAACAATGCCAATTTT 1682

Db 1582 GTGGGGTGAACCAAGAAAAACAAGCTTACACCCATATCATCTTCAAGAATGCCAATTTT 1641
Qy 1683 TACATTTACATGG--CATTCAGAGAACTAAT--CAGGGTCAAGATAATAGACGGTTCA 1738
Db 1642 TACATTTACATGGGCAATTTCCAGAGAACTAATTCAGGGTCCAAAGATAATAGACGGTTCC 1701
Qy 1739 TCAAT 1743
Db 1702 NCCAT 1706
RESULT 6
US-10-002-050-9
; Sequence 9, Application US/10002050
; Publication No. US20030032095A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corine
; APPLICANT: Yang, Meljia
; APPLICANT: Boldog, Ferenc
; APPLICANT: Herrmann, John
; TITLE OF INVENTION: No. US20030032095A1el Nucleic Acid Sequences Encoding Human Se
; FILE REFERENCE: 15966-554 Cura-54 CON-514
; CURRENT APPLICATION NUMBER: US/10/002,050
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 09/604,286
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/140,584
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1508
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (226)..(1458)
; NAME/KEY: variation
; LOCATION: (1)..(1508)
; OTHER INFORMATION: N may be any nucleotide
US-10-002-050-9
Query Match 33.3%; Score 1182.6; DB 9; Length 1508;
Best Local Similarity 98.4%; Pred. No. 0;
Matches 1228; Conservative 0; Mismatches 10; Indels 10; Gaps 3;
Qy 506 ACTCTTCTTGGATCCCTCGTGGAAACTACATAGAATCTAATCGTGATGACTGCACCGTGT 565
Db 230 ACTCTTCTTGGATCCCTCGTGGAAACTACATAGAATCTAATCGTGATGACTGCACCGTGT 289
Qy 566 CTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGCTATGCTTCTTTGAGTACCAGTATG 625
Db 290 CTTTGATCTATGCTGTGCACCTTTAAGAAGTCAGGCTATGCTTCTTTGAGTACCAGTATG 349
Qy 626 TCGACAACAACATCTTCTTTGAGTCTTTATTTCAAAATGATCAGTGCCAGGAGATGGACA 685
Db 350 TCGACAACAACATCTTCTTTGAGTCTTTATTTCAAAATGATCAGTGCCAGGAGATGGACA 409
Qy 686 CCACCACTCACAAGTGGGTAAAACTTTACAGACAATGGAGAATGGGGCTCTCATTTCTGTA 745
Db 410 CCACCACTCACAAGTGGGTAAAACTTTACAGACAATGGAGAATGGGGCTCTCATTTCTGTA 469
Qy 746 TCGTGAATTCAGGCACAAAACATCTACTTGGAGAACTACAGCATCTTATGGGTCTTA 805
Db 470 TCGTGAATTCAGGCACAAAACATCTACTTGGAGAACTACAGCATCTTATGGGTCTTA 529
Qy 806 AGCGGCTCAAGCCTGTGCTGGTAAAAAATATCACAAATGAAGGGGTGGCGTACACATCAG 865
Db 530 AGCGGCTCAAGCCTGTGCTGGTAAAAAATATCACAAATGAAGGGGTGGCGTACACATCAG 589
Qy 866 AATGTTTTCTTCCATGCAAGCCAGGCACATTCAGCAACAACAGGTTCTCATTTCAACTGCCAGG 925

```
Db 590 AATGTTTCTCTGCAAGCCAGGCACATTCAGCAACAAACAGGTTTCATCAACTGCCAGG 649
Qy 926 TGTGTCACAGAACACCTATTCTGAGAAGAGGAGCCAAAGAAATGTATAAGGTGAAGAGC 985
Db 650 TGTGTCACAGAACACCTATTCTGAGAAGAGGAGCCAAAGAAATGTATAAGGTGAAGAGC 709
Qy 986 ACTCTCAATTTTC-----AGGATCCAGTGTAGTACAGAGCGCCCTCCCTGTACCACAA 1039
Db 710 ACTCTCAATTTTCAGAGGAAGGATCCAGTGTAGTACAGAGCGCCCTCCCTGTACCACAA 769
Qy 1040 AAGACTATTTCCAGATCCATCTCCATGTGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1099
Db 770 AAGACTATTTCCAGATCCATCTCCATGTGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 829
Qy 1100 ACTGGATAGAGCCCAAAATCTCCGGGAGGATCTCCAGAGATCTTACAGATGCTTACAGAT 1159
Db 830 ACTGGATAGAGCCCAAAATCTCCGGGAGGATCTCCAGAGATCTTACAGATGCTTACAGAT 889
Qy 1160 CTGGAGAGAGGAGGATCTCCGGGAGGATCTCCAGAGATCTTACAGATGCTTACAGAT 1219
Db 890 CTGGAGAGAGGAGGATCTCCGGGAGGATCTCCAGAGATCTTACAGATGCTTACAGAT 849
Qy 1220 CTGGAGAGAGGAGGATCTCCGGGAGGATCTCCAGAGATCTTACAGATGCTTACAGAT 1279
Db 950 CTGGAGAGAGGAGGATCTCCGGGAGGATCTCCAGAGATCTTACAGATGCTTACAGAT 1009
Qy 1280 CAGCAGGAGAGGAGGATCTCCGGGAGGATCTCCAGAGATCTTACAGATGCTTACAGAT 1339
Db 1010 CAGCAGGAGAGGAGGATCTCCGGGAGGATCTCCAGAGATCTTACAGATGCTTACAGAT 1069
Qy 1340 ACATGAAACTTCTGCTTCAATGTGGGAATTTCAAGTGGGATGGAATGATGTTGGG 1399
Db 1070 ACATGAAACTTCTGCTTCAATGTGGGAATTTCAAGTGGGATGGAATGATGTTGGG 1129
Qy 1400 AGGTGGCTGGAGATCATATCCAGAGTGGGCTGGAGGATCTGCAATGATGATGATGAT 1459
Db 1130 AGGTGGCTGGAGATCATATCCAGAGTGGGCTGGAGGATCTGCAATGATGATGATGAT 1189
Qy 1460 TAACTTGCATATCCAGGATTTAAACCAACCAACATCTATGAGTGGGATGGAATGATGAT 1519
Db 1190 TAACTTGCATATCCAGGATTTAAACCAACCAACATCTATGAGTGGGATGGAATGATGAT 1249
Qy 1520 AACTAGGAGAGATACATTTGCTTTGAGACCTCTGTTGAGACCTGCTGTTGATGATGAT 1579
Db 1250 AACTAGGAGAGATACATTTGCTTTGAGACCTCTGTTGAGACCTGCTGTTGATGATGAT 1309
Qy 1580 TCATGGTGATATTAAGAAAAGTACAAATGTTGAGATGTTGAGATGTTGAGATGTTGAGAT 1639
Db 1310 TCATGGTGATATTAAGAAAAGTACAAATGTTGAGATGTTGAGATGTTGAGATGTTGAGAT 1369
Qy 1640 AAAAAAGAGTATACACCATATCATCTTCAAGAAATGCAAACTTTTACATTTACATGGG--C 1697
Db 1370 AAAAAAGAGTATACACCATATCATCTTCAAGAAATGCAAACTTTTACATTTACATGGG--C 1429
Qy 1698 ATTCAGAGAACTAAT--CAGGAGTCAAGATATAGAGGATGATGATGATGATGATGAT 1743
Db 1430 TTCCAGAGAACTAATTTACAGGTTCAAGATATAGAGGATGATGATGATGATGATGAT 1477
```

RESULT 7
US-10-002-304-9
; Sequence 9, Application US/10002304
; Publication No. US20030036185A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corine
; APPLICANT: Yang, Meijia
; APPLICANT: Boldog, Ferenc
; APPLICANT: Herrmann, John
; TITLE OF INVENTION: Polynucleotides and polypeptides encoded thereby
; FILE REFERENCE: 15966-554 Cura-54 CON-58

```
; CURRENT APPLICATION NUMBER: US/10/002.304  
; CURRENT FILING DATE: 2001-11-02  
; PRIOR APPLICATION NUMBER: 09/604.286  
; PRIOR FILING DATE: 2000-06-22  
; PRIOR APPLICATION NUMBER: 60/140.584  
; PRIOR FILING DATE: 1999-06-23  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 1508  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (226)..(1458)  
; NAME/KEY: variation  
; LOCATION: (1)..(1508)  
; OTHER INFORMATION: N may be any nucleotide  
US-10-002-304-9
```

```
Query Match 33.3%; Score 1182.6; DB 9; Length 1508;  
Best Local Similarity 98.4%; Pred. No. 0;  
Matches 1228; Conservative 0; Mismatches 10; Indels 10; Gaps 3;  
Qy 506 ACTCTTCTTTGGATCCCTCGTGGAACTACATAGAATCTAATCGTGATGACTGCACGGTGT 565  
Db 230 ACTCTTCTTTGGATCCCTCGTGGAACTACATAGAATCTAATCGTGATGACTGCACGGTGT 289  
Qy 566 CTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGCTATCTCTTTTGGTGGTACCATG 625  
Db 290 CTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGCTATCTCTTTTGGTGGTACCATG 349  
Qy 626 TCGACAACAACATCTCTTTTGGTGGTAACTTACAGCAATGAGAAATGGGCTCTCATTTCTG 685  
Db 350 TCGACAACAACATCTCTTTTGGTGGTAACTTACAGCAATGAGAAATGGGCTCTCATTTCTG 409  
Qy 686 CCACCACTGACAAGTGGTAACTTACAGCAATGAGAAATGGGCTCTCATTTCTG 745  
Db 410 CCACCACTGACAAGTGGTAACTTACAGCAATGAGAAATGGGCTCTCATTTCTG 469  
Qy 746 TCGTGAATTCAGGCACAAACATCTCTACTGGAGAACTACAGCATCTTATGGGTTCTA 805  
Db 470 TCGTGAATTCAGGCACAAACATCTCTACTGGAGAACTACAGCATCTTATGGGTTCTA 529  
Qy 806 AGCGGTCAAGCTGTGCTGTTAAATAATCACAATTTGAAGGGTGGCTACACATCAG 865  
Db 530 AGCGGTCAAGCTGTGCTGTTAAATAATCACAATTTGAAGGGTGGCTACACATCAG 589  
Qy 866 AATGTTTTCTTTGCAAGCCAGGCACATTCAGCAACAAACAGGTTTCATTCAGTCCAGG 925  
Db 590 AATGTTTTCTTTGCAAGCCAGGCACATTCAGCAACAAACAGGTTTCATTCAGTCCAGG 649  
Qy 926 TGTGTCACAGAACACCTATTCTGAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 985  
Db 650 TGTGTCACAGAACACCTATTCTGAGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 709  
Qy 986 ACTCTCAATTTTC-----AGGATCCAGTGTAGTACAGAGCGCCCTCCCTGTACCACAA 1039  
Db 710 ACTCTCAATTTTCAGAGGAAGGATCCAGTGTAGTACAGAGCGCCCTCCCTGTACCACAA 769  
Qy 1040 AAGACTATTTCCAGATCCATCTCCATGTGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1099  
Db 770 AAGACTATTTCCAGATCCATCTCCATGTGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 829  
Qy 1100 ACTGGATAGAGCCCAAAATCTCCGGGAGGATCTCCAGAGATCTTACAGATGCTTACAGAT 1159  
Db 830 ACTGGATAGAGCCCAAAATCTCCGGGAGGATCTCCAGAGATCTTACAGATGCTTACAGAT 889  
Qy 1160 CTGGAGAGAGGAGGATCTCCGGGAGGATCTCCAGAGATCTTACAGATGCTTACAGAT 1219  
Db 890 CTGGAGAGAGGAGGATCTCCGGGAGGATCTCCAGAGATCTTACAGATGCTTACAGAT 949  
Qy 1220 CTGGAGAGAGGAGGATCTCCGGGAGGATCTCCAGAGATCTTACAGATGCTTACAGAT 1279
```

|||||
Db 950 CTTGGCATTCCCTCTCTCGAACAATTTTCAGATGGAAACCAAGAATGTAGACCATTGTC 1009
|||||
Qy 1280 CAGCAGGAACGAGGCTGCACCTTGGCTTTGAATATAAATGGTGGAAATGCTTCTTGCGCA 1339
|||||
Db 1010 CAGCAGGAACGAGGCTGCACCTTGGCTTTGAATATAAATGGTGGAAATGCTTCTTGCGCA 1069
|||||
Qy 1340 ACATGAAACTTCCCTGCTTCAATGTTGGGAATTTCAAAGTGGCATGGAATGAATGTTGGG 1399
|||||
Db 1070 ACATGAAACTTCCCTGCTTCAATGTTGGGAATTTCAAAGTGGCATGGAATGAATGTTGGG 1129
|||||
Qy 1400 AGGTGCTGGAGATCATATCCAGATGGGGCTGGAGGTTCTGACAAATGATTAACCTGATCT 1459
|||||
Db 1130 AGGTGCTGGAGATCATATCCAGATGGGGCTGGAGGTTCTGACAAATGATTAACCTGATCT 1189
|||||
Qy 1460 TAAACTTGCATATCCAGGATTTAAACCAACCAACATCTATGACTGGAGCCAGGGTTCTG 1519
|||||
Db 1190 TAACTTGCATATCCAGGATTTAAACCAACCAACATCTATGACTGGAGCCAGGGTTCTG 1249
|||||
Qy 1520 AACTAGGAAGAATAACATTTGTCTTTGAGACCTCTGTTCAGCTGACTGTGTTGTACT 1579
|||||
Db 1250 AACTAGGAAGAATAACATTTGTCTTTGAGACCTCTGTTCAGCTGACTGTGTTGTACT 1309
|||||
Qy 1580 TCATGGTGAATTAATAGAAAAAGTACAAATGTGGTAGAATCGTGGGTGGAACCAAG 1639
|||||
Db 1310 TCATGGTGAATTAATAGAAAAAGTACAAATGTGGTAGAATCGTGGGTGGAACCAAG 1369
|||||
Qy 1640 AAAACAAGCTTACACCCATATCATCTTCAAGAAATGCAACTTTTACATTTACATGGG--C 1697
|||||
Db 1370 AAAACAAGCTTACACCCATATCATCTTCAAGAAATGCAACTTTTACATTTACATGGGCA 1429
|||||
Qy 1698 ATTCCAGAGAATAAT--CAGGGTCAAGATAATAGACGGTTTCATCAAT 1743
|||||
Db 1430 TTCCAGAGAATAATTTACGGGTCCAAGATAATAGACGGTTCCNCCAT 1477
|||||

RESULT 8

US-10-003-152-9
; Sequence 9, Application US/10003152
; Patent No. US20020151494A1
; GENERAL INFORMATION:
; APPLICANT: Shimkets, Richard
; APPLICANT: Fernandes, Elma
; APPLICANT: Vernet, Corine
; APPLICANT: Yang, Meljia
; APPLICANT: Boldog, Ferenc
; APPLICANT: Herrmann, John
; TITLE OF INVENTION: No. US20020151494A1el Amino Acid Sequences for Human Semaphorin-1
; FILE REFERENCE: 15966-554 Cura-54 CON-S12
; CURRENT APPLICATION NUMBER: US/10/003,152
; PRIOR FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 09/604,286
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/140,584
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1508
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (226)..(1458)
; NAME/KEY: variation
; LOCATION: (1)..(1508)
; OTHER INFORMATION: N may be any nucleotide
US-10-003-152-9

Query Match 33.3%; Score 1182.6; DB 12; Length 1508;
Best Local Similarity 98.4%; Pred. No. 0;
Matches 1228; Conservative 0; Mismatches 10; Indels 10; Gaps 3;

Qy 506 ACTCTCTTTGGATCCCTCGTGGAAACTACATAGAATCTAATCGTGATGACTGCACGGTCT 565
|||||
Db 230 ACTCTCTTTGGATCCCTCGTGGAAACTACATAGAATCTAATCGTGATGACTGCACGGTCT 289
|||||
Qy 566 CTTTGTATCTATGCTGTGCACCTTAAGAAGTCAGGCTATGCTTCTTTGAGTACCACTATG 625
|||||
Db 290 CTTTGTATCTATGCTGTGCACCTTAAGAAGTCAGGCTATGCTTCTTTGAGTACCACTATG 349
|||||
Qy 626 TCGACAACAACATCTTTTGTAGTCTTTTATTCAAAATGATCAGTCCAGGAGATGCACA 685
|||||
Db 350 TCGACNACNACATCTTCTTTGAGTCTTTATTCAAAATGATCAGTCCAGGAGATGCACA 409
|||||
Qy 686 CCACCACTGACAAGTGGGTAAACCTTACAGACAATGAGAATGGGCTCTCATTTCTGTAA 745
|||||
Db 410 CCACCACACACAAGTGGGTAAACCTTACAGACAATGAGAATGGGCTCTCATTTCTGTAA 469
|||||
Qy 746 TCGTGAATTCAGGCACAACATCTCTACTGGNAGACTACAGGCATCTTATGGGTTCTTA 805
|||||
Db 470 TCGTGAATTCAGGCACAACATCTCTACTGGNAGACTACAGGCATCTTATGGGTTCTTA 529
|||||
Qy 806 AGCGGCTCAAGCTGTGCTGGTAAAAAATATCACAAATTCGAAGGGTGGGTACACATCAG 865
|||||
Db 530 AGCGGCTCAAGCTGTGCTGGTAAAAAATATCACAAATTCGAAGGGTGGGTACACATCAG 589
|||||
Qy 866 AATGTTTTCTTTGCAAGCCAGGCACATTCAGCAACAACAGGTTCTTCACTGCCAGG 925
|||||
Db 590 AATGTTTTCTTTGCAAGCCAGGCACATTCAGCAACAACAGGTTCTTCACTGCCAGG 649
|||||
Qy 926 TCTGTCCCAAGAACACCTTATTTCTGAGAAAGGAGCCAAAGAAATGTATAAGGTGTAAAGACG 985
|||||
Db 650 TGTGTCCCAAGAACACCTTATTTCTGAGAAAGGAGCCAAAGAAATGTATAAGGTGTAAAGACG 709
|||||
Qy 986 ACTCTCAATTTTTC-----AGGATTCAGTGAGTACAGAGGCCCTCCCTGTACCACAA 1039
|||||
Db 710 ACTCTCAATTTTTCAGAGGAAGGATCCAGTGAGTACAGAGGCCCTCCCTGTACCACAA 769
|||||
Qy 1040 AAGACTATTTCCAGATCCATCTCCATGTGATGAAGAAGAAAGACACAGATAATGTACA 1099
|||||
Db 770 AAGACTATTTCCAGATCCATCTCCATGTGATGAAGAAGAAAGACACAGATAATGTACA 829
|||||
Qy 1100 AGTGGATAGAGCCCAAAATCTCCGGGAGGATCTCAGATGCTATAGATTGCCCCCTT 1159
|||||
Db 830 AGTGGATAGAGCCCAAAATCTCCGGGAGGATCTCAGATGCTATAGATTGCCCCCTT 889
|||||
Qy 1160 CTGGAGAGAAGAGGATTTGTCGCCCTTGCAACCCCTGGATTTTATAAATGGATCATCTT 1219
|||||
Db 890 CTGGAGAGAAGAGGATTTGTCGCCCTTGCAACCCCTGGATTTTATAAATGGATCATCTT 949
|||||
Qy 1220 CTTTGCATCCCTGTCCTCTCGAACAATTTTCAGATGGGAACCAAGAAATGTAGACCAATGTC 1279
|||||
Db 950 CTTTGCATCCCTGTCCTCTCGAACAATTTTCAGATGGGAACCAAGAAATGTAGACCAATGTC 1009
|||||
Qy 1280 CAGCAGGAACGAGGCTGCACCTTGGCTTTGAATATAAATGGTGGAAATGCTTCTTGCGCA 1339
|||||
Db 1010 CAGCAGGAACGAGGCTGCACCTTGGCTTTGAATATAAATGGTGGAAATGCTTCTTGCGCA 1069
|||||
Qy 1340 ACATGAAACTTCCCTGCTTCAATGTTGGGAATTTCAAAGTGGCATGGAATGAATGTTGGG 1399
|||||
Db 1070 ACATGAAACTTCCCTGCTTCAATGTTGGGAATTTCAAAGTGGCATGGAATGAATGTTGGG 1129
|||||
Qy 1400 AGGTGCTGGAGATCATATCCAGATGGGGCTGGAGGTTCTGACAAATGATTAACCTGATCT 1459
|||||
Db 1130 AGGTGCTGGAGATCATATCCAGATGGGGCTGGAGGTTCTGACAAATGATTAACCTGATCT 1189
|||||
Qy 1460 TAAACTTGCATATCCAGGATTTAAACCAACCAACATCTATGACTGGAGCCAGGGTTCTG 1519
|||||
Db 1190 TAACTTGCATATCCAGGATTTAAACCAACCAACATCTATGACTGGAGCCAGGGTTCTG 1249
|||||
Qy 1520 AACTAGGAAGAATAACATTTGTCTTTGAGACCTCTGTTCAGCTGACTGTGTTGTACT 1579
|||||
Db 1250 AACTAGGAAGAATAACATTTGTCTTTGAGACCTCTGTTCAGCTGACTGTGTTGTACT 1309
|||||
Qy 1580 TCATGGTGAATTAATAGAAAAAGTACAAATGTGCTAGAAATCGTGGGTGGGAACCAAG 1639

```
|||||
Db 1310 TCATGGTGAATTAATAGAAAAAGTACAAATCTGTAGAAATCGTGGGTGGAACCAAG 1369
QY 1640 AAAAAAGCTTACACCCATATCATCTTCAAGAAATGCAACTTTTACATTTACATGGG--C 1697
|||||
Db 1370 AAAAAAGCTTACACCCATATCATCTTCAAGAAATGCAACTTTTACATTTACATGGGGA 1429
QY 1698 ATTCCAGAGAACTAAT--CAGGCTCAAGATAATAGACGGTTCATCAAT 1743
|||||
Db 1430 TTCCAGAGAACTAATTCAGGCTCCAAGATAATAGACGGTTCNCCAT 1477

RESULT 9
US-10-028-072-37
; Sequence 37, Application US/10028072
; Publication No. US2003000431A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang
; TITLE OF INVENTION:
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/028,072
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059836
; PRIOR FILING DATE: 1997-09-24
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/062285
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/062287
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/062814
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/062816
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063045
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063082

; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/063127
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063327
; PRIOR FILING DATE: 1997-10-27
; PRIOR APPLICATION NUMBER: 60/063329
; PRIOR FILING DATE: 1997-10-27
; PRIOR APPLICATION NUMBER: 60/063550
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063561
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063704
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/063733
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/063735
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/063738
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/063755
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/064248
; PRIOR FILING DATE: 1997-11-03
; PRIOR APPLICATION NUMBER: 60/064809
; PRIOR FILING DATE: 1997-11-07
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065846
; PRIOR FILING DATE: 1997-11-17
; PRIOR APPLICATION NUMBER: 60/066364
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/066453
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/066511
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/069212
; PRIOR FILING DATE: 1997-12-11
; PRIOR APPLICATION NUMBER: 60/069278
; PRIOR FILING DATE: 1997-12-11
; PRIOR APPLICATION NUMBER: 60/069334
; PRIOR FILING DATE: 1997-12-11
; PRIOR APPLICATION NUMBER: 60/069694
; PRIOR FILING DATE: 1997-12-16
; PRIOR APPLICATION NUMBER: 60/072320
; PRIOR FILING DATE: 1998-01-23
; PRIOR APPLICATION NUMBER: 60/073612
; PRIOR FILING DATE: 1998-02-04
; PRIOR APPLICATION NUMBER: 60/074086
; PRIOR FILING DATE: 1998-02-09
; PRIOR APPLICATION NUMBER: 60/074092
; PRIOR FILING DATE: 1998-02-09
; PRIOR APPLICATION NUMBER: 60/077791
; PRIOR FILING DATE: 1998-03-12
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079294
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: 60/079663
; PRIOR FILING DATE: 1998-02-27
; PRIOR APPLICATION NUMBER: 60/079728
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/080165
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/081203
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081229
; PRIOR FILING DATE: 1998-04-09
; PRIOR APPLICATION NUMBER: 60/081695
; PRIOR FILING DATE: 1998-04-14
; PRIOR APPLICATION NUMBER: 60/081817
; PRIOR FILING DATE: 1998-04-15
```


Db 1338 CTCCAATGG---CTCAGACTGTACCCGCTGCCCTCCAGGACTGAACCTGCTGTGGGATT 1394
Qy 1308 TGAATATAAATGGTGAATGCTTCCTGGCAACATGAAACTTCTCTGCTCAATGTTGG 1367
Db 1395 TGAATACAAATGGTGAACACGCTGCCCAACAACATGAAAGACGCTCTCAGTGGAT 1454
Qy 1368 GAATTCAAAGTCGGATGGAATGAATGGTGGAGGTGGCTGGAGATCATATCCAGATGG 1427
Db 1455 CAATTCGAGTACAGGCGATGACAGGCTGGAGGTGGCTGGTATCATATTCACACAGC 1514
Qy 1428 GGCTGGAGGTCTGACATGATTAACCTGATCTTAACCTGATATCCAGAGTTTAAACC 1487
Db 1515 TGCTGGAGCTCAGACAATGACTTCTCATGATTTCTACTCTGTTGTGCCAGGATTTAGACC 1574
Qy 1488 ACCAATCTATGACTGGAGCCAC---GGTCTCAACTAGGAAGAAATAAATTTGCTTT 1544
Db 1575 TCCGAGTGGGTGATGGCAGACACAGAAATAAGAGGTGGCCAGAATCACATTTGCTTT 1634
Qy 1545 TGAGACCTCTCTGCTCAGCTGACTGTGTTTGTACTTTCATGGTGGATATTAATAGAAAAG 1604
Db 1635 TGAGACCTCTCTGCTGCACTGTGAGCTCTACTTTCATGGTGGTGTCAATTTCTAGGAC 1694
Qy 1605 TACAATGTGTAGATCGTGGGTGGAAACCAAGAAAGAAAGCTTACACCATATCAT 1664
Db 1695 CAACACTCTGTGGAGACGTGGAAGGTTTCCAAAGGCAAAACAGTCTATACCTACATCAT 1754
Qy 1665 CTTCAAGATCAACTTTTACATTTTACATGGGCATTCAGAGAACTAATCAGGGTCAAGA 1724
Db 1755 TGAGGAGAACACTACACAGAGTCTCACCCTGGGCCCTTCCAGAGGACACTTTTTCATGAGGC 1814
Qy 1725 TAATAGACGGTTCATCAATGACATGATGGTGAAGATTTATCTATCAGAGCCACTTAATGCGAGT 1784
Db 1815 AAGCAGGAGTACACCAATGAGCTTGCCAGATCTACTCCATCAATGTCCACCAATGTTAT 1874
Qy 1785 TGATGGGTGGGTCTCTATGCGGTGCGCTGTGCGCTGCTGTGAAACAGTGGGTTCATC 1844
Db 1875 GAATGCGTGGGTCTCTACTGCGCTGCGCTGTGCGCTGTGAAAGCTCTGTATGGGTCTCTC 1934
Qy 1845 GTGTGCTCCCTGCCCTCCAGGCACTACATTCAGAAAGAAACCAACAGGTGCAAGGAATG 1904
Db 1935 CTCGACCTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1994
Qy 1905 TCCACTGACACCTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1964
Db 1995 CCCCCCTAACACAATCTGAAAGCCCAACAGCTTATGCTGCTGCTGCTGCTGCTGCTGCTG 2054
Qy 1965 CCGGCTGGAGTAAACAATCAGGACCACTTGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2024
Db 2055 TGTCCAGGAGCCAGAAACAAGATCCACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2114
Qy 2025 CCATGAAAAAGAAATCAGATTTTGCACTATGACTTTTACCAACCTCAGCAGTGTGGGCTC 2084
Db 2115 ACCCAACTCCAAACAGAGCTTTCAACTACACTTCTCCGCTTGGCAACACCGCTCAC 2174
Qy 2085 ATTAATGAATGGCCCCAGCTTCACTCCAAAGGAACAAATACTTCCATTTCTTCAATAT 2144
Db 2175 TCTTGTGGAGGCCAAGCTTCACTTCCAAAGGTTGAAATACTTCCATCACTTTACCCT 2234
Qy 2145 CAGTTTATGTGGCATGAGGGGAGAGATGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2204
Db 2235 CAGTCTCTGTGGAACACAGGTTAGSAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2294
Qy 2205 TACAGTAAAGAAATAGTGGCAGGTCAGATGATTAACAATAATTTGTTAGGGCAATTTGT 2264
Db 2295 CCGATTCCTGAGGTGAGTCAGG-----TTCTCCAATCTATCACAGCCTACGT 2345
Qy 2265 ATGCCAGTCAACAATTTATCTTCTGAAAGTAAGGTTTCCGAGAGCCCTTATCATCACA 2324
Db 2346 CTGCCAGGAGTATCATATCCCCCAGAGGTGACAGGCTACAGGCCGGGTTTCTCTCACA 2405
Qy 2325 ATCCATCTTCTGGCAGATACATTCATAGGATGACAGTTGAAACCAATTTGAAAAATAT 2384

Db 2406 GCCTGTCCAGCCTTGTGCTGATCGACTTAATTTGGGTGACAACAGATATGACTCTGGATGAAT 2465
Qy 2385 TAATATAAAGAAAGATATGTTCCAGATTCCCAAGCAAAATACAGATGTGCATTTCTT 2444
Db 2466 CACCTCCCAAGCTGAACCTTTTCCACCTGGAGTCTTGGGAATACCGGACGTGATCTTCTT 2525
Qy 2445 TTATAAGTCTTCTACAGCAACAACATCTTGTATTAATGGCCGATCAACTGCTGTGAAAT 2504
Db 2526 TTATAGGTCCAATGATGTGACCCAGTCTGCAGTCTTGGAGATCAACCAACATCCGGGT 2585
Qy 2505 GAGGTGTAATCTTACTTAATCTGGAGCAGGAGTATTTAGTCTCCCAAGAGTGGCCAGC 2564
Db 2586 CAGGTGCACTGCCAGAAAACCTGCTCCCTGGAAGTTGCTGCTGCCAGAACTGCTCAGA 2645
Qy 2565 AGGTACCTGTGATGGGTGTACGTTCTATTCTCTGGGAGAGTGTGAAGCTTGGCCCTCT 2624
Db 2646 TGGACCTGTGATGCTGCAACTTCCACTTCTGCTGGAGAGCGGGTCTGCTGGCCCT 2705
Qy 2625 GTGTACGGAGCATGACTTCCATGAGATTTGAGGAGCCCTGCAAGAGAGATTTTCAGAGAAC 2684
Db 2706 CTGCTCAGTGGCTGACTACCATGCTATCTGCTCAGCAGCTGTGTGGCTGGGATCCAGAGAC 2765
Qy 2685 CTTGTATGTGTGGAATGAACCTAATGATGATGATTAAGAGATTTCTTTGCCTGAGAAAA 2744
Db 2766 TACTTACGTGTGNCAGAGAACCCAGCTATGCTGTGCTGGCATTTCTGCTGCTGAGCAGAG 2825
Qy 2745 GTTGGCAACCTGTGAACGGTTGACTTTTGGCTGAAGGTGGGAGCGGTGTGGGAGCTTT 2804
Db 2826 AGTCACCATCTGCAAAACCATAGATTTCTGCTGAAAGTGGGCATCTCTGCGGACCTG 2885
Qy 2805 TACTCCGCTTTTGTGGTGGCTCTGACCTGCTTCTTGGAAAAAGAAATCAAAAACTGGA 2864
Db 2886 TACTGCCATCTGCTCACCCTGCTTGCACCTGCTTCTTGGAAAAAGAAATCAAAAACTAGA 2945
Qy 2865 ATACAAATATTCAGATTTAGTAATGACGACTTAACCTCAAAAGAGTGTGAACCTCCCGCTGC 2924
Db 2946 GTACAAGTACTCCAAGCTGGTGTGATGAATCTACTCTCAAGGACTGTGACCTCCAGCAGC 3005
Qy 2925 ACAGAGTGTGCTATCATGGAAGAGAGATTAATGAAGAGGAAAGTTGTATATTTCCAATAA 2984
Db 3006 TGACAGCTCGGCATCATGGAAGCGGAGATGTAGAGCAGCCTCATCTTTTACCAGCAA 3065
Qy 2985 ACAGTCACTACTAGGAAAACTCAAACTCTTTGGCAACCAAGGAAAAAGAACCATTTTGA 3044
Db 3066 GAAGTCACTTTTGGAGAGATCAAAATCAATTACCTCCAGAGGAGCTCTCTGATGGATTGA 3125
Qy 3045 ATCTGTTCAACTCAAAACCTC 3065
Db 3126 CTCAGTGGCGCTGAAGACATC 3146

RESULT 10

US-10-121-049-37

; Sequence 37, Application US/10121049

; Publication No. US2003002239A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;; FILE REFERENCE: P33303C17
;; CURRENT APPLICATION NUMBER: US/10/121,049
;; CURRENT FILING DATE: 2002-04-12

;; Prior Application removed - See File Wrapper or Palm

;; NUMBER OF SEQ ID NOS: 550

;; SEQ ID NO 37

;; LENGTH: 3501

;; TYPE: DNA

;; ORGANISM: Homo Sapien

;; FEATURE:

;; NAME/KEY: unsure

;; LOCATION: 2762, 2778

;; OTHER INFORMATION: unknown base

us-10-121-049-37

Query Match 27.8%; Score 987.8; DB 9; Length 3501;

Best Local Similarity 60.5%; Pred. No. 1.7e-264;

Matches 1755; Conservative 0; Mismatches 1119; Indels 27; Gaps 7;

QY 171 GCTTCCTCCTTGCAGGAGAAAGATTATACACTTTTGAATATACGGAATGTGATAGCAGTGG 230
DB 267 GCTTCATGCTGCAAGAGTCTGAGTACCACATATGATACAGGGCTGTGACAGCAGGG 326
QY 231 CTCAGGTGGAGAGTTGCCATTCCAAATTTCTGCAGTGGACTGCTCTGGCCTGCCTGACCC 290
DB 327 TTCAGGTGGAGGTGCGCGTCCGATACCCCGGCGCTGTGCACAGCCTGTCTGACCC 386
QY 291 AGTCAGAGCAAGATGACATTTCTCCTCTGCTTCTGAGAGATCTAGAAATGAAGAA 350
DB 387 CGTCAAGGCGCAGGTGCTCTCTCTCCGCAACGCGGGGAGTTTCTGGATATGAAGA 446
QY 351 CCAGGTATGCAATAGTGGTGAAGGCACTTATCTCTGGGCGAGTGCATCAAAATTGCA 410
DB 447 CCAGTCATGTAAGCCATGCGCTGAGGCGCTACTCTCCCTCGGCACAGGATTCGGTTGA 506
QY 411 TGAATGGGATGAATTCGCGGAGGATTTCTTACATTCGAATTCATGACACACTGTGT 470
DB 507 TGAGTGGGATGAGCTGCCCATGGCTTTCCAGCCCTCTCAGGCAACATGGAGCTGGATGA 566
QY 471 GGCCCTCTCTGACAGCAGGCGCAGAGGCTGTAAACACTCTCTTGGATCCCTCGTGGAA 530
DB 567 CAGTGTCTGAGTCCA---CGGGGAATGTACTTCGTCGAAGTGGTTCCTCCGGGGCGCA 623
QY 531 CTACATAGAAATTAATGCTGATGCTGACGCTGTCTTTGATCTATGCTGTGCACTTAA 590
DB 624 CTACATCGGCTCCACACAGCAGCAATGACGCCACACTGATGTACGCGTCAACCTGAA 683
QY 591 GAAGTCAGGCTATGCTCTTTTGGTACCAGTATGTCGACACACATCTTCTTTGAGTT 650
DB 684 GCAATCTGGCACCCTTAAGTTCGAATGATCTTCCAGACTCCAGACTCATCTTTGAGTT 743
QY 651 CTTTATTTCAAAATGATCAGTGGCAGGAGATGGACACCACCTGACAGTGGGTAAACT 710
DB 744 TTTCTGTCAGATGACAGTGGCCGCAATCCAGATGACTC---CAGGTGATGAAGAC 800
QY 711 TACAGAAATGGAGATGGGGCTCTCATCTGTAATGCTGAAATCAGGCACAAACATACT 770
DB 801 CACAGAGAA---AGGATGGAAATTCACAGTGTGGAGCTAAATTCAGGCAATATGTCCT 857
QY 771 CTACTGGAGAACTACAGGATCCTTATGGGTTCTAAGCGGCTCAAGCCTGTGCTGGTAA 830
DB 858 CTATTGGAGAACCAAGCCTTCTCAGTATGGACCAAAAGTACCACAGCCTGTGCTGGTGA 917
QY 831 AAATATCAAAATTAAGGGGTGGCTTACACATCAGAAATGTTTCTTCTTCAAGCCAGGAC 890
DB 918 AAACATGGCAATACAGGGGTGGCTTACATCTCAGATGCTTCCCTGCAACACTGGGAC 977
QY 891 ATTACACAAACACAGGTTTCAATCACTGCGAGGTGTGTCAGAAACACATATCTGA 950
DB 978 GTATGCAGACACAGGCGCTCTCTTTCTGCAAACTTTGCCAGCAGCACTCTTATTCAA 1037

QY 951 GAAAGAGCAAAAGAAATGTATAGGTGTAA---AGACGACTCTCAATTTTCAGGATCCAG 1007
DB 1038 TAAAGGAGAAACTTTTGGCCACAGTGTGACCTTGACAATATCTCAGAAAGGATCTTTC 1097
QY 1008 TGAGTGTACAGAGCGCCCTCCTGTACCACAAAAGACTATTTTCCAGATCCATCTCCATG 1067
DB 1098 TTTCTCTAAGCTGGCGCCAGCTTGCACACAAAGATTTATTTCTACACACACACGGCCTG 1157
QY 1068 TGATGAGAAGAAAGACACACAGATATGTACAAAGTGGATAGAGCCCAAAATCTGCCGGGA 1127
DB 1158 CGATGCCAAACGGAGAGACACAACTCATGTACAAATGGGCAAGCGGAAATCTGTAGCGA 1217
QY 1128 GGATCTCACAGATGCTATTAGATTTGCCCTTCTGAGAGAAAGAGGATTTGCCGCTTG 1187
DB 1218 GGACCTTGAGGGGCGAGTGAAGCTGCCCTCTCTGTGTGAAGACCCACTGCCACCCCTG 1277
QY 1188 CAACCTTGGAATTTTATTAACAAATGGATCATCTTTGCCATCCCTGTCTCTCTCCGAAAT 1247
DB 1278 CAACCCAGGCTTTCTCAAAACCAACACAGCACCCTGCCAGCCCTGCCCATATGTTCTCTA 1337
QY 1248 TTCAGATGGAACCAAGAGATGTAGACCATGTCCACAGGAGGAGGAGCCTGCACTTGGCT 1307
DB 1338 CTCCAATGG---CTCAGACTGTACCCGCTGCCCTGCAAGGACTGAACCTGTCTGTGGGAT 1394
QY 1308 TGAATATAAATGCTGGAATGCTCTTCTCTGGCAACATGAAACCTTCTCTCAATGTTGG 1367
DB 1395 TGAATACAAATGCTGGAACACGCTGCCCAACAAACATGGAACGACCGTCTCTCAGTGGGAT 1454
QY 1368 GAAATCAAAGTCGATGGAATGAATGATGTTGGAGGTGGCTGGAGATCATATCCAGAGTGG 1427
DB 1455 CAACTTCGAGTACAAGGCGATGACAGGCTGGAGGTGGCTGGTGTATCATATTTACACAGC 1514
QY 1428 GCTCGAGGCTCTGACAAATGATTAACCTTAACTTGCATATCCAGAGATTTAAACC 1487
DB 1515 TCTCGAGCCTCAGACAATGACTTTCATGATCTCCTACTCTGGTGTGGCAGGATTTAGACC 1574
QY 1488 ACCAATCTATGACTGGAGCCAC---GGGTTCTGAACTAGGAAGAAATAAATTTGTCT 1544
DB 1575 TCCGAGTGGGTGATGGGACACACAGAGAAATGAAGGTGGCCAGAAATCACATTTGCTTT 1634
QY 1545 TGAGACCTCTCTGTCAGCTGACTGTGTTTGTACTTTCATGGTGGATATTAATAGAAAAG 1604
DB 1635 TGAGACCTCTCTGTTGAACTGTGAGCTGTACTTCTACTTCTGTTGGTGGTGTGAATTTAG 1694
QY 1605 TACAATGTGTTAGAAATGCTGGGTGGAAACCAAAACAAAGCTTACACCATATCAT 1664
DB 1695 CAACACTCTCTGGAGAGCTGGAAGGTTCANAGGCAACACAGTCTTATACCTACATCT 1754
QY 1665 CTTTCAAGAAATGCAACTTTTACATTTTACATGGGCAATTCAGAGAACTAATCAGGGTCA 1724
DB 1755 TGAGGAGAACTTACACAGAGCTTCCACCTGGGCTTCCAGAGGAGCCACTTTTCATGAG 1814
QY 1725 TAATGAGCGTTTCAATCAATGATGCTGAAGATTTATTTCTATCACAGCCACTAATGCACT 1784
DB 1815 AAGCAGAGTGTACCAATGACCTTGGCAAGATCTACTTCCATCAATGTCCACCAATGTTAT 1874
QY 1785 TGATGGGTGGCGCTCTCATGCGCTGCTGTGCTTCTGAAACAGTCCGGTTCATC 1844
DB 1875 GAATGGCGTGGCGCTCTACTGCGCTGCTGTGCTTCTGAAAGCTCTGATGTGGGCTCCTC 1934
QY 1845 GTGTGTCCTCCCTTCCAGGCGCACTACATTTGAGAAAGAAACCAACACAGTGCAGGAATG 1904
DB 1935 CTGCACTCTTGTCTGCTGCTGTTTACTATATTTGACGAGATTCAGGAACCTGCCACTCCT 1994
QY 1905 TCCACCTCACACCTTACCTGCTTATCATACATCAGTCTATGGAAGAGGCTTGTATTCCATG 1964
DB 1995 CCCCCTTAACCAATTTGAAAGCCCAAGCCTTATGTTGGTGTCCAGGCGCTGTGTGCCCTG 2034
QY 1965 CGGCGCTGGGAGTAAAAACAATCAGGACCAATTCGCTTATAGTACTGCTTTTCTTCTA 2024
DB 2055 TGGTCCAGGAGCAAGAAACAAAGATCCACTCTCTGTGTGCTACATGATGATCCACTTCTC 2114
QY 2025 CCATGAAAAGAAATCAGATTTTGCACATATGACTTTTAGCAACCTCAGCAGTGTGGGCTC 2084

```
Db 2115 ACGRACACTCTCAACACAGGACTTTCAACTACAACTTCTCCGCTTTGGCAACACCGTCTAC 2174
QY 2085 ATTAATGAATGCCCGCAGCTTCACCTCCAAAGGACAAATACTTCCATTTCTTCAATAT 2144
Db 2175 TCTTGCTGGAGGCCAAGCTTCACTTCCAAGGGTTGAATCTTCCATCACTTTACCT 2234
QY 2145 CAGTTTATGTGGCATGAGGGAAGAAGTCTCTGTACCAACAATATAACAGACTT 2204
Db 2235 CAGTCTCTGTGAACACAGGGTAGGAATGTCTGTGCACCGAATGTCACTGACT 2294
QY 2205 TACAGTAAAGAAATAGTGGCAGGTCAGATGATTACACAAATTTGGTAGGGCAATTTGT 2264
Db 2295 CCGGATTCCTGAGGGTAGTCAAGG-----TTCCTCAAATCTATCACAGCCTACGT 2345
QY 2265 ATGCCAGTCAACAATATTCTCTCTGAAAGTAAGGGTTCCAGCAGCGCTTATCATCACA 2324
Db 2346 CTGGCAGGAGCATCATATCCCCCAGAGGTGACAGGCTACAAAGCGGGGTTTCTCTACA 2405
QY 2325 ATCCATCTCTGGCAGATACATTCATAGGAGTCACAGTTGAAACACACATTTGAAAAATAT 2384
Db 2406 GCCTGTCAACCTTGTGATCGACTTATTTGGGTGACACAGATATCACTCTGGATGGAT 2465
QY 2385 TAATATAAAGAGATATGTTCCAGTTCCTCAAGTCCCAAGCCAAATACCAGATGTGCATTTCT 2444
Db 2466 CACCTCCCGAGCTGAATTTTCCACCTGGAGTCTTGGGAATACCGGAGCTGATCTTCTT 2525
QY 2445 TTATAGTCTTACAGCAACAACATCTTGTATTAATGCCGATCAACCTGTGTAATAT 2504
Db 2526 TTATAGTCTTACAGCAACAACATCTTGTATTAATGCCGATCAACCTGTGTAATAT 2585
QY 2505 GAGTGTAACTCTTAAATCTGGACAGGAGTGATTTTCACTCCCGCAGCAAGTGCACAGC 2564
Db 2586 CAGTGTCACTCCACAGAAACTGTCCCTGGAAGTTTGTCTGCCAGGAACGTGCTCACA 2645
QY 2565 AGTACCTGTGATGGGTGACGTTCTATTTCTGTGGGAGAGTGCTGGAAGCTTGGCCTCT 2624
Db 2646 TGGGACCTGTGATGGGTGCAACTTCCACTTCTGTGGGAGAGCGGCTGTGTCGCCGCT 2705
QY 2625 GTCTAGGAGCATGACTTCCATGAGATTTGAGGAGGCTCCAGAGAGGATTCAGGAAC 2684
Db 2706 CTGCTAGTGGCTGTACCATGCTATCTGTCAGCAGCTGTGTGGTGGGATCCAGANGAC 2765
QY 2685 CTGTGTATGTGGAATCAACTAAATGGTGCAATTAAGGAATTTCTTGCCTGAGAAAAA 2744
Db 2766 TACTTACGTGTGNCAGAGACCCCAAGCTATCTCTGTGGCAATTTCTCTGCTGAGCAG 2825
QY 2745 GTTGGCAACCTGTGAACGGTTGACTTTTGGCTGAAGGTGGGAGCGGTGTGGGAGCTTT 2804
Db 2826 AGTCACCATCTGCAAAACCATAGATTTCTGGCTGAAAGTGGGCATCTCTGCAGGCACCTG 2885
QY 2805 TACTGCGGTTTCTGTGGCTCTGACCTGCTACTTCTGGAAGAAATCAAAAACTGGA 2864
Db 2886 TACTGCCATCTCTGCTCACCCTTTGACCTGTGACTTCTCAAGGAGCTGTGACCTGCCAGC 2945
QY 2865 ATACAAATATTCAGTTAGTATGACGACTAACTCAAAAGAGTGTGAATCCCGGCTGC 2924
Db 2946 GTCAAGTACTCAAGCTGTGTGAATGCTACTCTCAAGGAGCTGTGACCTGCCAGCAGC 3005
QY 2925 AGACAGTTGTCTATCATGAAGAGAGAGATATGAAGAGGAAGTTGTATATTTCCATTA 2984
Db 3006 TGACAGCTGCGCCATCATGGAAGCGAGGATGTAGAGGAGGCTCATCTTTTACCAGCAA 3065
QY 2985 ACAGTCTACTTAGGAAACTCAAACTTTTGGCAACCAAGGAAAAAGAACACCATTTTGA 3044
Db 3066 GAAGTCACTTTTGGGAAGATCAATCATTTTACCTCCAAGAGGACTCTGATGGATTGA 3125
QY 3045 ATCTGTTCAACTGAAACCTC 3065
Db 3126 CTCAGTCCGCTGAAGACATC 3146
```

RESULT 11

```
US-10-123-904-37
: Sequence 37, Application US/10123904
: Publication No. US20030022328A1
: GENERAL INFORMATION:
: APPLICANT: Baker, Kevin P.
: APPLICANT: Beresini, Maureen
: APPLICANT: DeForge, Laura
: APPLICANT: Desnoyers, Luc
: APPLICANT: Filvaroff, Ellen
: APPLICANT: Gao, Wei-Qiang
: APPLICANT: Gerritsen, Mary B.
: APPLICANT: Goddard, Audrey
: APPLICANT: Godowski, Paul J.
: APPLICANT: Gurney, Austin L.
: APPLICANT: Sherwood, Steven
: APPLICANT: Smith, Victoria
: APPLICANT: Stewart, Timothy A.
: APPLICANT: Tumas, Daniel
: APPLICANT: Watanabe, Colin K
: APPLICANT: Wood, William
: APPLICANT: Zhang, Zemin
: TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
: TITLE OF INVENTION: ACIDS ENCODING THE SAME
: FILE REFERENCE: P330R1C34
: CURRENT APPLICATION NUMBER: US/10/123,904
: CURRENT FILING DATE: 2002-04-16
: Prior Application removed - See File Wrapper or Palm
: NUMBER OF SEQ ID NOS: 550
: SEQ ID NO 37
: LENGTH: 3501
: TYPE: DNA
: ORGANISM: Homo Sapien
: FEATURE:
: NAME/KEY: unsure
: LOCATION: 2762, 2778
: OTHER INFORMATION: unknown base
US-10-123-904-37
```

```
Query Match 27.8%; Score 987.8; DB 9; Length 3501;
Best Local Similarity 60.5%; Pred. No. 1.7e-264;
Matches 1755; Conservative 0; Mismatches 1119; Indels 27; Gaps 7;

QY 171 GCTTCTCTCTGCCAGGAGAAAGATTACATTTTGAATATACGGAATGTGTAGCAGTGG 230
Db 267 GCTTCATGCTGCAAGAGTCTGTAGTACCACTATGAGTACACGCGGTGTGACAGCAGGG 326
QY 231 CTCAGGTGGAGAGTTGCCATTCCAAATTCGCGTGGACTCTCTGCGCTCCCTGACCC 290
Db 327 TTCCAGGTGGAGGGTTCGCGTCCGATACCCCGGCGCTGTGCACAGCCTCTCTGACCC 386
QY 291 AGTGAGAGCAAGAAATGCACATTTCTCTCTGTCTTCTTGAGAGATATCTAGAAATGAAGA 350
Db 387 CGTCAAGGCGCAGGAGTGTCTCTCTGCAACGCCGGGAGTTCTGTGATATGAAGA 446
QY 351 CCAGGTATCCAGTAAGTGTGGTGAAGGCACCTATTTCCTTGGGCGAGTGGCATCAATTTGA 410
Db 447 CCAGTCATGTAAAGCCATGCGCTGAGGCGCGCTACTCCCTCGCACAGGCAATTCGGTTTGA 506
QY 411 TGAATGGGATGAATTCGCGGAGAGATTTCCTAACATCGCAACATTCATGGACACTGTGGT 470
Db 507 TGAGTGGGATGAGTGCCTCCCATGGCTTTGCCAGCCTCTCAGCCAACATGGAGCTGGATGA 566
QY 471 GGGCCCTTCTGACAGAGCCAGCGGTGTGAACAACTCTTCTTGGATCCCTCGTGGAAA 530
Db 567 CAGTGTCTGTAGTCCA---CCGGGAACGTACTCTGCTCAAGTGGTTCCCGGGCGCA 623
QY 531 CTACATAGTAATCTAATCTGTGATGACGAGGTGTCTTTGATCTATGCTGTGCACCTTAA 590
Db 624 CTACATCGCTCCACACAGGAGCAATGACAGCCACTGATGTAGCGCGTCAACCTGAA 683
QY 591 GAAGTCAGGCTATGCTCTCTTTTGTAGTACCAGTATGTCGACAAACAACTCTTCTTGTAGTT 650
Db 684 GCAATCTGCGACCGTTAACTTCTGAATCTACTATCCAGACTCCAGCATCATCTTGTAGTT 743
```

Qy	651	CTTTATTCAAATGATCAGTCCGACGAGATGGACACACCACCTGCAACTGGGTAAAAC	710
Db	744	TTTCGTTCAGAATGACCAAGTCCGACCCCAATGCAGATGACTC---CAGGTGGATGAAC	800
Qy	711	TACAGCAANTGGAGAAATGGGGCTCTCATTTCTGTAAATCGTGAATCAGGCACAACAT	770
Db	801	CACAGAGAA---AGGATGGGAATTCACACAGTGTGGAGCTTAAATTCGAGGCCAATAT	857
Qy	771	CTACTGGAGAACTACAGGCATCCTTATGGGTCTTAAGGGCGTCAAGCGCTGTCCTGGT	830
Db	858	CTATTGGAGACCACACGCCCTTCAGTATGGACCAAGTATCCCAAGCCTGTGTGGTGAG	917
Qy	831	AAATATCAAAATGAAGGGTGGCGTACACATCAGAAATGTTTCTTGTCAAGCCAGGCAC	890
Db	918	AAACATTTGCCATAACAGGGTGGCCTACACTTTCAGAAATCCTTCCCTCGCAAACTGG	977
Qy	891	ATTGAGCAACAAACAGGTTCAATTCACATGCCAGTGTGCTCCACAGAAACACTATTCT	950
Db	978	GTATGCGACAGCAGGGCTCTCTTTCTGCAAACTTTGCCCAGGCCAACTCTTATTCAA	1037
Qy	951	GAAGAGCCCAAGAAATGTATAAGGTGTA---AGACGACTCTCAATTTTCAGGATCCAG	1007
Db	1038	TAAAGGAGAACTTCTTGCCACCAGTGTACCTGCACAAATCTCAGAGAAAGGATCTTC	1097
Qy	1008	TGAGTGTACAGAGCGCCCTCCCTGTACACACAAGAACTATTTCAGATGCAATCTCCAT	1067
Db	1098	TTCTCTTAACGTGGGCCAGTGTGCACAGACAAAGATATTATTCTACACACACAGCGG	1157
Qy	1068	TGATGAAGAAGGAACACAGATAATGTACAAGTGGATAGAGCCCAAAATCTCCCGGGA	1127
Db	1158	CGATGCCAACGGAGAGACAACTCATGTACAAATGGCCAGCCGAAATCTGTACCGA	1217
Qy	1128	GGATCTCACAGATGCTATTAGATTGCCCCCTTCTGGAGAGAAGAGATTCTCGCCCTTG	1187
Db	1218	GGACCTTGGAGGGGCAGTAGAGCTGCCCTGCCCTGTGTGAAGACCCACTGCCACCC	1277
Qy	1188	CAACCCCTGGATTTTATAACAATGGATCATCTTCTTGCCATCCCTGTCTCTCTGGAAC	1247
Db	1278	CAOCCAGGCTTCTCAAAACCAACAACGACCTGCCAGCCCTGCCCATATGGTTCTCTA	1337
Qy	1248	TTCAGATGGAAACCAAGATGTAGACCATGTCCACAGGAAGGAGCCTGCACCTTGGCTT	1307
Db	1338	CTCCAATGG---CTCAGATGTACCCGCTGCCCTCGAGGACTGAACCTGCTGTGGGAT	1394
Qy	1308	TGAATATAAATGGTGGAAATGCTCTTCTGGCAACATGAAACTTCTCTCTCAATGTTGG	1367
Db	1395	TGAATACAAATGTTGGAACACGCTGCCACAACATGGAACGACCGTTCTCAGTGGGAT	1454
Qy	1368	GAATTCAACTGGCATGGAAATGAATGTTGGGAGTGGCTGGAGCATCATATCCAGAGTGG	1427
Db	1455	CAACTTCGAGTACAAAGGGATACAGAGCTGGGAGTGGCTGGTGTATCACATTATACAG	1514
Qy	1428	GGTGGAGGTTCTGACAAATGATTACTGTATCTTAACTTGCATATCCAGGNTTTAAACC	1487
Db	1515	TGCTGGAGCTCAGACAATGACTTCAATGATCTCACTCTGGTTGTGCCAGGATTTAGACC	1574
Qy	1488	ACCAACATATWAGACTGGAGCCAC---GGGTCTGAACTAGGAAGAATAACATTTGTCTT	1544
Db	1575	TCGCGAGTCGGTATGGCAGACACAGAAATAAGAGGTGCCCAGANTCACATTTGTCTT	1634
Qy	1545	TGAGACCTCTGTTACGCTGACTGTGTTTGTACTTTCATGTGATATTAATAGAAAAG	1604
Db	1635	TGAGACCTCTGTTCTGTGAATGTGAGCTCTACTTTCATGTGGTGTGAATTTCTAGGAC	1694
Qy	1605	TACAAATGTGGTAGAATCGTGGGGTGGAAACCAAGAAACAAAGCTTTCACCCCATAT	1664
Db	1695	CAACACTCCTGTGGAGACGTGGAAAGGTTCCAAAGGCAACAGTCTTATACCTACATCAT	1754
Qy	1665	CTTTCAGAAATGCAACTTTTACATTTACATGGGCATTCACAGAACTAATCAGGGTCAAG	1724
Db	1755	TGAGGAGAACTATACACAGCTTCACCTGGCCCTCCAGGAGCACTTTTCATCAGGC	1814

QY	1725	TAATAGACGGTTTCATCAATGACATGGTGAAGATTTATTTCTATCACAGCCACTAATGACGT	1784
DB	1815	AAGCAGGAATACACCAATACGCTTGCCAAGATCTACTCCATCAATGTCCACCAATGTTAT	1874
QY	1785	TGATGGGTGGCGTCCFCATGCGCTGCTGCGCCCTCGGTTCTTCAACAGTCGGGTTCATC	1844
DB	1875	GAATGGGTGGCCTCTTACTTGGCGTCCCTGTGCCCTAGAAGCCTCTGATGTGGGCTGCCTC	1934
QY	1845	GTGTGTCCTTGGCCTCCAGGCCACTACATTTAGAAAGAAACCAACACAGTTCGAAGGAATG	1904
DB	1935	CTGCACCTCTTGCTGCTGTTACTATATTGCGGAGATTCAGGAACCTGCCACCTCCTG	1994
QY	1905	TCCACCTTGACACCTACCTGTCCATACATCAGTCTATTGGCAAGAGAGCTGTGTTCATG	1964
DB	1995	CCCCCTTAACCAAAATTTCTGAAGCCCAACAGCCTTATGSGTCCACGGCCTGTGTGCCCTG	2054
QY	1965	CGGGCCTGGGAGTAAACCAATCAGGACCAATTCGGTTTTGCTATAGTACATGCTTTTTCCTA	2024
DB	2055	TGTTCCAGGGACCAAGCAACAACTCCACTCTCTGTGCTACAAATGATTTGCACCTTCTC	2114
QY	2025	CCATGAAAGAANAATCAGATTTTTCGACATATGACTTTTAGCAAGCTCAGCAGTGTGGGCTC	2084
DB	2115	ACCAACACTTCCAACAGGACTTTTCAACTACAATTTCTCCGCTTTGCCAACAACCCCTCAC	2174
QY	2085	ATTTAATGAATGGCCCGCAGCTTCACTTCCAAGCAACAAATACTTCCATTTCTCAATAT	2144
DB	2175	TCTTGTCTGGAGGCCAAGCTTCACTTCCAAGGGTTGAATACTTCCATCACCTTTACCT	2234
QY	2145	CAGTTTTATGTGGGCATGAGGGGAAGATGCTCTCTGTACCAACAAATAACAGACTT	2204
DB	2235	CAGTCTCTGTGGAACACGAGGTAGGAAATGTCTGTGCACCGCAATGTCACTGACCT	2294
QY	2205	TACAGTAAAGAAATAGTGGCAGGTCAGATGATACACAAATTTGCTAGGGGCATTTGT	2264
DB	2295	CCGGATTCTGAGGTTGAGTCAGGG-----TTCTCCAATCTATCACAGCCTACCT	2345
QY	2265	ATGCCAGTCAACAATTTATCTCTTGAAAGTAGGGTTTCCGAGCAGCCTTATCATCACA	2324
DB	2346	CTGCCAGGCACTCATCATCCCCCAGAGGTGACAGCTACAAGGCCGGGTTTCTCTCACA	2405
QY	2325	ATCCATCATTTGGCAGATACATTCATAGAGTTCACAGTTGAAACCCACATTTGAAATAAT	2384
DB	2406	GCCTGTACGCTTGTCTGATCGACTTATTTGGGTTGACACAGATATGACTCTGGATGGAA	2465
QY	2385	TAATATAAAGAAATATGTTCCCACTTCCAACAGCCAAATACCAAGATGTGCATTTCTT	2444
DB	2466	CACCTCCCCAGCTGAACATTTTCCACCTGGAGTCTTGGGCAATACCGGACGTGATCTCTT	2525
QY	2445	TTATAGTCTTCTACAGCAACACATCTTTGTATTAATGGCCGATACACTGCTGTGAAAT	2504
DB	2526	TTATAGTCCAATGATGTACCCAGTCTCTGCAATTTGGGAGATCAACCAACCATCCGCT	2585
QY	2505	GAGGTGTAATCTTACTAAATCTGGAGCAGAGCTGATTTTCAGTCCCCACGAAGTGCCACG	2564
DB	2586	CAGGTGCAGTCCACAGAAACTCTCCCTGGAAGTTTGTGCTGCCAGGAACGTGCTCAGA	2645
QY	2565	AGTACCTGTGATGGGTGACGTTCTATTTCTGTGGGAGAGTGTGAAGCTTGGCCTGT	2624
DB	2646	TGGGACCTGTGATGGTGCACACTTCCACTTCTGTGGGAGCGGGCTGCTTTGCCGCT	2705
QY	2625	GTGTACGGAGCATCTTCCATGAGATTCAGGAGCCTTGCAGAGAGGATTTTCAGGAAC	2684
DB	2706	CTGCTCAGTGGCTGACTACCATGCTTATCGTCAGCAGCTGTGTGGCTGGATCCAGANGAC	2765
QY	2685	CTTGTATGTGGGAATGAACCTAAATGGTGCATTAAGGAATTTCTTTTGCCTGAGAAAA	2744
DB	2766	TACTTTACGTGTGNCGAGAACCACCAAGCTATGCTCTGGTGGCATTTCTCTGCTGTGACAG	2825
QY	2745	GTGTGCACACTCTGAACGGTTGACTTTTGGCTGAAGTGGGAGCCGCTGTGGGAGCTTTT	2804
DB	2826	AGTCACCATCTGCAAAACCAATAGATTTCTGGCTGAAGTGGGCATCTCTGCAAGCACCTG	2885
QY	2805	TACTCCCGTTTTGCTGTGGTCTGACCTGTACTCTGTGMAAAAAAGAAATCAAAAACTGGA	2864


```
QY 1428 GCGTGGAGGTTCTGACAATGATTACCTGATCTTAAACTTTGCATATCCCGAGGATTTAAACC 1487
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1515 TGCTGGAGCCTCAGACAATGACTTTCATGATCTCACTCTGGTGTGCCAGGATTTAGACC 1574
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1488 ACCAATCTATGACTGGAGCCAC---GGGTCTGACTAGGAGAAATAACATTTGCTCTT 1544
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1575 TCCGAGTCGGTGATGGCAGACAGAGAAATAAGAGGTGGCCAGATCACAATTTGCTCTT 1634
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1545 TGAGACCCCTCTGTTACGCTGACTGTGTTGTTACTTTCATGGTGGATATTAATAGAAAAG 1604
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1635 TGAGACCCCTCTGTTCTGTGAACCTGTGAGCTTACTTTCATGGTGGTGTGAATTTCTAGGAC 1694
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1605 TACAAATGTGGTAGAATCTGGGGTGGAAACCAAGAAAGAAAGCTTACACCCATATCAT 1664
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1695 CAACACTCTCTGTGGAGACGTGGAAAGTTCCAAAGGCAAAACAGTCTTATACCTACATCAT 1754
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1665 CTTCAAGATGCAACTTTTACATTTTACATTTGCGCATTCAGAGAACTAATCAGGGTCAAGA 1724
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1755 TGAGGAGAACACTTACCACGAGGTTTACCTGGGCTTCCAGAGGACCACTTTTTCATGAAGC 1814
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1725 TAATAGACGTTTCAATGACATGGTGAAGATTTATTTCTATCACAGCCACTTAATGCACT 1784
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1815 AAGCAGAAAGTACACCAATGACGTTGCCAAGATCTACTCCATCAATGTCAACCAATGTTAT 1874
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1785 TGATGGGTGGGCTCTCATATGCCGTGCCCTGTGCCCTCGGTTCTGAACAGTCGGGTTCAFC 1844
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1875 GAATGGCTGGGCTCTACTTGGCGCTGCTGCTGCTAGAGGCTCTGATGTGGGCTCCTC 1934
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1845 GTGTGTCCCTGCGCTTCAGGCGCACTTACATTTGAGAAAGAAACCAACAGTGTCAAGGAATG 1904
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1935 CTGCACTCTTGTCTGCTGCTGTTACTATATTGACCGAGATTCAGGAACCTGCCACTCCTG 1994
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1905 TCCACTGACACCTACTCTGCCATACATCAGTCTATGCGAAGAGGCTGTGATTTCCATG 1964
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1995 CCCCCCTAACACAATTTCAAAAGCCCAACAGCCTTATGTGTCCAGGCGTGTGTGCCCTG 2054
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1965 CGGCGCTGGAGTAAACAATCAGACCACTTCGGTTCCTATAGTACTGCTTTTCTTA 2024
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2055 TGGTCAGGAGCCAGAACACAGATCCACTCTCTGTGCTACAATGATTCACCTTCTC 2114
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2025 CCATGAAAAGAAATCAGATTTTGCACTTATGACTTTTAGCAACCTCAGCAGTGTGGGCTC 2084
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2115 ACSCAACACTCCAAACAGGACTTCAACTACACATTCCTCGCTTTGGCAACACCGCTCAC 2174
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2085 ATTAATGAATGGCCCCAGCTTCACCTCCAAAGGACAAATACTTCATTTCTTCAATAT 2144
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2175 TCTTGTGGAGGGCCAAAGCTTCACTTCCAAAGGGTTGAAATATCTTCCATCACTTTACCT 2234
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2145 CAGTTTATGTGGCATGAGGGAAGAGATGGCTCTCTGTACCAACAATATTAACAGACTT 2204
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2235 CAGTCTCTGGAACACAGGCTAGGAAATGTCTGTGTGCACCGAACAATGTCACCTGACCT 2294
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2205 TACAGTAAAGAAATAGTGGCAGGCTCAGATGATTACACAAATTTGGTAGGGGCATTTGT 2264
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2295 CCGGATTCCTGAGGTGTAGTCAGG-----TTCCTCAAATCTATCACAGCCTACCT 2345
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2265 ATGCCAGTCAACAATATTTCCTTCTGAAAGTAAAGGTTTCCAGCAGCCTTATCATCACA 2324
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2346 CTGCCAGGAGCATCATCCCCCAGGCTGACAGGCTACAAGCGCGGGTTTCTCTCACA 2405
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2325 ATCCATCTTCTGGCAGATACATTCATAGGAGTTCAGAGTTGAAAGCACATTTGAAAATAT 2384
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2406 GCCTGTCACTTGTGTGCTGCTGCTTATTTGGGTGACACAGATATGACTCTGGATGGAT 2465
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2385 TAATATAAGAGATATGTTCCAGTTTCCACAGCCAAATACCAGATGTGCTCATTTCTT 2444
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2466 CACCTCCCAAGCTGAACTTTTCCACCTGGAGTCTTGGGAATACCGGAGCTGATCTTCTT 2525
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2445 TTATAAGTCTTCTACAGCAACAACATCTTGTATTAATGGCCGATCAACTGCTGTGAAAT 2504
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2526 TTATAGTCCATGATGTGACCCAGCTCTGCTGCTGAGATCAACCACTCCGGT 2585
```

```
QY 2505 GAGGTGTAATCTACTAAATCTGGAGCAGGAGTGAATTTAGTCCCGCAGCAAGTCCCGCAGC 2564
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2586 CAGGTGCAGTCCACAGAAAACCTGTCCTTGGAAATTTGCTGCTGCCAGAACGCTCTCAGA 2645
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2565 AGGTACCTGTGATGGGTGCTAGCTTCTATTTTCTTGTGGAGAGTGTGTAAGCTTCCCTCT 2624
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2646 TGGAGACCTGTGATGGCTGCAACTTCCACTTCTCTGTGGAGAGCGGCTGCTTCCCGCT 2705
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2625 GTGTACGAGGAGTCACTTTCATGAGATTTGAGGAGCGCTTGAAGAGAGAGATTTTCAGGAAC 2684
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2706 CTGCTCAGTGGCTGACTACCAATGCTATCTGTCAGCAGCTGTGTGGCTGGGATCCAGAGAC 2765
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2685 CTTGTATGTGTGGAATGAACCTAAATGGTGCATTTAAAGGAATTTCTTTTGGCTGAGAAAAA 2744
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2766 TACTACCTGTGNCAGAAACCAAGCTATGCTCTGCTGGTGGCATTTCTCTGCTGAGCAGAG 2825
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2745 GTTGGCAACCTGTGAAAGGTTGACTTTTGGCTGAAGGTGGAGCCGCTGTGGGAGCTTT 2804
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2826 AGTCACCATCTGCAAAACCATAGATTTCTTGGCTGAAAGTGGGCATCTCTGCGAGGCACCTG 2885
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2805 TACTGCCCTTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2864
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2886 TACTGCCATCTCTGCTCACCCTGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2945
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2865 ATACAAATATTTCAAGTTAGTAATGACGAGCTAATCAAAAGAGTGTGAACTCCCGGCTGC 2924
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2946 GTACAAGTACTTCCAAGCTGGTGTGATCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3005
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2925 AGACAGTTGTCTATCATGGAAGGAGAGATTAATGAAGAGAACTGTATATTTCCAATAA 2984
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3006 TGACAGCTGCGCCATCATGGAAGCGGAGGATGTAGAGGACGACCTCATCTTTACCAGCAA 3065
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2985 ACAGTCACTACTAGGAAACTCAAACTTTTGGCAACCAAGAAAGAAAGAACCACTTTTGA 3044
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3066 GAAGTCACTTTTGGGAAGATCAAAATCATTTACCTCAAGAGGACTCTCTGATGATTTGA 3125
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3045 ATCTGTCAACTGAANAACCTC 3065
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3126 CTCAGTGGCGCTGAAGACATC 3146
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 13

US-10-175-746-37

; Sequence 37, Application US/10175746

; Publication No. US20030027270A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

; APPLICANT: DeForge, Laura

; APPLICANT: Desnoyers, Luc

; APPLICANT: Flivaroff, Ellen

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Gerritsen, Mary E.

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Sherwood, Steven

; APPLICANT: Smith, Victoria

; APPLICANT: Stewart, Timothy A.

; APPLICANT: Tumas, Daniel

; APPLICANT: Watanabe, Colin K

; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; FILE OF INVENTION: ACIDS ENCODING THE SAME

; CURRENT APPLICATION NUMBER: US/10/175,746

; CURRENT FILING DATE: 2002-06-19

; Prior Application removed - See File Wrapper or PalM

; NUMBER OF SEQ ID NOS: 550

; SEQ ID NO 37

; LENGTH: 3501

; TYPE: DNA


```
QY 2205 TACAGTAAAGAAATAGTGGCAGGGTCAGATGATTACACAAATTTGGTAGGGCATTGT 2264
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2295 CCGATTCTCTGAGGGTCAGTCAGG-----TTCTCCAAATCTATACACACCTACGT 2345
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2265 ATGCCAGTCAACAATATTCCTCTGAAAGTAGGGTTCCGAGCAGCCTTATCATCACA 2324
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2346 CTCCAGGAGTATCATATCCCGCAGAGGTGACAGGCTACAAGCGGGGTTTCTCTACA 2405
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2325 ATCCATCATCTGGCAGATACATTCATAGGATCAGAGTTGAACACACATTTGAAATAT 2384
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2406 GCCTGTGACGCTTGTGATCGACTTATGGGGTGACACAGATATGACTCTGGATGGAAT 2465
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2385 TAATATAAAGAGATATGTTCCAGTTCACAAAGCCAAATACCAGATGTGCTATTTCTT 2444
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2466 CACCTCCCGACGTGAATTTTCCACCTGGAGTCTTTGGGAATACCGGAGTGATCTCTT 2525
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2445 TTATAGTCTCTACAGCACACATCTGTATTAATGCGCGATCAACTGCTGTGAAAT 2504
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2526 TTATAGTCCAAATGATGTGACCCAGTCTCTGAGTCTCTGGGAGATCAACCACTCCGCT 2585
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2505 GAGGTGTATCTTACTAAATCTGGAGCAGGAGTGATTTTCAGTCCCGCAGCAAGTCCCGC 2564
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2586 CAGGTGAGTCCACAGAAACTCTCCCTGAGTTTCTGCTGCCAGAACGTGCTCACA 2645
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2565 AGTACCTGTGATGGGTGATGCTTCTATTTCTGTTGGGAGAGTGTGAAGCTTGCCCTCT 2624
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2646 TGGGACCTGTGATGGGTGCAACTTCCACTTCTCTGTGGGAGAGCGGCTGCTGCGCGCT 2705
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2625 GTGTAGGAGCATGACTTCCATGAGATGAGGAGGCTTCAAGAGAGGATTTTCAGGAAC 2684
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2706 CTGCTCAGTGGCTGACTACCATGCTATGTCAGCAGCTGTGTGGTGGGATCCAGANGAC 2765
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2685 CTTGTATGTGGAATGAACCTAAATGGTGCAATTAAGGAATTTCTTCCCTGAGAAAAA 2744
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2766 TACTTACGTGTGNCGAGAACCCAAAGCTATGCTCTGTGTGGCATTTCTCTGCTGAGCAG 2825
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2745 GTTGGCAACTGTGAACGGTTGACATTTTGGCTGAGAGTGGGAGCGGTGTGGGACCTTT 2804
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2826 AGTCACCATCTGCAAAACCATAGATTTTGGCTGAAAGTGGGCATCTGTGAGGACCTTG 2885
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2805 TACTGCGGTTTGGCTGTGGCTGTGACCTGTCTACTTTCTGGAAGAAAGAAATCAAAACTGA 2864
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2886 TACTGCACTCTCTCACCCTCTTGACCTGTACTTTTGGAAAGAAAGAAATCAAAACTAGA 2945
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2865 ATACAATATTCAGTGTAGTATGACGACTAATCAAAAGAGTGTGAAGTCCCGGCTGC 2924
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2946 GTACAAGTACTCCAGCTGGTGATGAATGCTACTCTCAAGAGGACTGTGACCTGCCAGCAGC 3005
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2925 AGACAGTTGTCTATCATGGAAGGAGAAAGATATGAAGAGGAGTGTGATATTTCCAATAA 2984
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3006 TGACAGCTGCGCCATCATGGAAGGCGAGGATGTAGAGGACGACCTCATCTTTACCAGCAA 3065
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2985 ACAGTCACTACTAGGAAATCAAAATCTTTGGCAACCAAGGAAAGAAAGAACCATTTTGA 3044
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3066 GAAGTCACTTTTGGGAGAGTCAAAATCATTTTACCTCCAAGAGGACTCTCTGATGGATTGA 3125
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3045 ATCTGTTCACTCAAAACCTC 3065
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 3126 CTCAGTGGCCGCTGAAGACATC 3146
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 14

```
US-10-176-918-37
; Sequence 37, Application US/10176918
; Publication No. US2003002725A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
```

```
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: WOOD, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P330R1C382
; CURRENT APPLICATION NUMBER: US/10/176,918
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 37
; LENGTH: 3501
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 2762, 2778
; OTHER INFORMATION: unknown base
US-10-176-918-37
```

Query Match 27.8%; Score 987.8; DB 9; Length 3501;

Best Local Similarity 60.5%; Pred. No. 1.7e-264;

Matches 1755; Conservative 0; Mismatches 1119; Indels 27; Gaps 7;

```
QY 171 GCTTCTCTCTTGGCCAGGAGAAAGATTATCACTTTTGAATATACGGAATCTGTAGCAGTGG 230
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 267 GCTTCATGCTGCAAGAGTCTGAGTACCACCTATGAGTACAGGGGTGTGACAGCACGGG 326
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 231 CTCAGAGTGGAGAGTTGCCATTTCCAAATTTGCGAGTGAGCTGCTGCGCTGCGCTGACCC 290
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 327 TTTCCAGTGGAGGGTCCGCGTCCCGCATACCCCGGCGCTGTGCCACAGCTGTCTGACCC 386
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 291 AGTGAGAGCGCAAGAAATGCATTTCTCTCTGTCTTCTCGAGAGTATCTAGAAATGAACAA 350
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 387 CGTGAGGCGCACCGAGTCTCTTCTCTCTGCGACGCCGGGAGTTTCTGGATATGAAGGA 446
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 351 CCAGTATGTCAGTAAAGTGTGTGAAGGACCACTATTTCTTGGGAGTGGCATCAAAATTTGA 410
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 447 CCAGTATGTAAGCCATCGGCTGAGGGCGGCTACTTCCCTCGGCACAGGCAATCGGTTTGA 506
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 411 TGAATGGGATGAATTTGCCGCGAGGATTTTCAACATCGCAACATTCATGGACACTGTGCT 470
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 507 TGAGTGGGATGAGTGTGCGCCCATGGCTTTTGGCAGGCTCTCAGCCAACATGGAGCTGGATGA 566
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 471 GGGCCCTTCTGACAGCAGCGCAGAGCGGTGTAAACAACTTTCTTGGATCCCTCGTGGAAA 530
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 567 CAGTCTGCTGAGTCCA--CGGGAACTGTACTTCTGTCCAAGTGGGTTCGCCCGGGCGCA 623
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 531 CTACATAGAATCTAAATCTGTATGACTGCAACCGTGTCTTTTGTATCTATGCTGTGCACTTAA 590
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 624 CTACATCCCTCCAACACGAGCAATGACAGCCACACTGATGTACGCGGTCAACCTGAA 683
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 591 GAAATCAGGCTATGCTTTTGTAGTACAGTATGTCGACAAACAACTCTTCTTTGAGTT 650
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 684 GCAATCTGCCACCGTTAACTTCGAAATCTACTATCCAGACTCCAGCACTCATCTTTTGTAGTT 743
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 651 CTTTATTCAAAATGATGTCGCCAGGAGATGGACACCACTGACAAAGTGGGTGAAACT 710
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 744 TTTCTGTTCAAGATGACCACTGCCAGCCCAATGCAGATGACTC---CAGGTGATGAAGAC 800
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 711 TACACACAATGAGAAATGGGCTCTCATTTCTGTAAATCAGGCACAAACATCACT 770
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 801 CACAGAGAA---AGGATGGGAATTTCCACAGTGTGGAGCTTAAATCCAGGCAATAATGCTCT 857
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 771 CTACTGGAGAACTACAGGCATCTTATGGTGTCTTAAGCGCGTCAAGCCTGTGCTGGTAA 830
Db      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db 858 CTATTGGAGAACACAGAGCTTCTCAGTATGACCAAAAGTACCCAGAGCTGTGCTGGTGGAG 917
Qy 831 AAATATCAACAATTGAAGGGGTGGGTACACATCAGAAATGTTTTCTTGCAGAGCCAGGCAC 890
Db 918 AAACATGGCCATACAGAGGGGTGGGTACACATTCAGAAATGTTTTCTTCCCTGCAAAACCTGGCAC 977
Qy 891 ATTACAGCAACAACAGAGTTTCAATCACTGACAGTGTGTCAGAGCAACACCTATTCTGA 950
Db 978 GTATCGAGACAAGCAGGGCTCTCTTTCTGCAAACTTTGGCCAGGCCAACTCTTATTTCAA 1037
Qy 951 GAAAGAGCCAAAGAAATGTAAGGTGTA---AGACGACTCTCAATTTTCAGGATCCAG 1007
Db 1038 TAAAGGAGAACTTCTTGCCACAGGTGTGACCCCTGACAAATCTCAGAGAAAGGATCTTC 1097
Qy 1008 TGAGTGTACAGAGCCGCTTCTGTACCAACAAGAGCTATTTCCAGAGTCCATCTCCATG 1067
Db 1098 TTCTGTAACTGTGCGCCAGCTTGCACAGCAAAAGATTATTTCTACACACACAGCCGCTG 1157
Qy 1068 TGNATGAAGAGAAACACACAGATATGTACAGTGGATAGAGCCCAAAATCTGCCGGA 1127
Db 1158 CGATGCCAACCGAGAGACAACTCATGTACAAATGGGCCAAGCCGCAAAATCTGTAGCGA 1217
Qy 1128 GGATCTCACAGATGCTATTAGATTGGCCCTTCTGGAGAGAAAGGATTGTCCGCTTG 1187
Db 1218 GGACTTGGAGGGCAGTGAAGTGGCTGCTTGTGTGAAGACCCACTGCTCCACCCCTG 1277
Qy 1188 CAACCTGGATTATTAACAATGGATCATCTTCTGCCATCCCTGCTCTCTCTGGGAACAT 1247
Db 1278 CAACCCAGGCTTCTTCAAAACCAACACAGCACCTGCCAGCCCTGCCATATGGTTCTTA 1337
Qy 1248 TTCAGATGGNACCAAGAAATGTAGACCATGTCCAGCAGAACGGAGCTGCACTTGGCT 1307
Db 1338 CTCCAATGG---CTCAGACTGTACCCGCTGCCCTGCGAGGACTGAACCTGCTGTGGGAT 1394
Qy 1308 TGAATATAATGTGGAAATGCTCTCTGGCAACATGAAACTTCTGCTTCAATTTGG 1367
Db 1395 TGAATACAATGTGGNACACGCTGCCACAAACATGGAACGACCGTTCTCAGTGGAT 1454
Qy 1368 GAATTCAGAGTGGATGAATGATGGTTGGGAGGTGGCTGGAGATCATATCCAGAGTGG 1427
Db 1455 CAACTTCAGTAAAGGGCATACAGCTGGGAGGTGGCTGGATCACAATTTACACAGC 1514
Qy 1428 GGCTGGAGGTCTGACAAATGATTACCTGATCTTAAACTTGCATATCCAGGATTTAAAC 1487
Db 1515 TGCTGGAGCCTCAGACAACTGACTTCATGATTTCTCACTCTGGTTGTGCCAGGATTTAGAC 1574
Qy 1488 ACCAATCTATGACTGGAGCCAC---GGTTCTGAACTAGGAAGAAATACATTTGTCTT 1544
Db 1575 TCCGACGTCGGTATGGCAGACACAGAGAATAAGAGGTGGCCAGAAATCACAATTTGCTT 1634
Qy 1545 TGAGACCTCTGTTCAGCTGACTGCTTTTGTACTTTCATGCTGGATTAATAGAAAAG 1604
Db 1635 TGAGACCTCTGTGTTGTGAACCTGTGAGCTCTACTTCTGTTGGGTGGAATTTCTAGGAC 1694
Qy 1605 TACAAATGTGTAGATCGTGGGTGGAACCAAGAAAACAAAGCTTTACACCCATATCAT 1664
Db 1695 CAACACTCTGTGGAGCTGGAAAGTTTCCAAAGGCAACAGTCTCTATACCTATCAT 1754
Qy 1665 CTTCAAGAAATGCAACTTTTACATTTTACATGGCAATTCAGAGAACTAATCAGGCTCAAGA 1724
Db 1755 TGAGGAGAACTACTACACAGGCTTCACTCTGGGCTTCCAGAGGACCACTTTTTCATGAGC 1814
Qy 1725 TAATAGAGGTTTCATCAATGACATGGTGAAGATTTATTTCTATCACGCCACTAATCAGT 1784
Db 1815 AAGCAGGAAGTACACCACTAGCTTGGCAAGATCTACTTCCATCAATGTACCAATGTTAT 1874
Qy 1785 TGATGGGTGGGCTCTCATCCGCTGCTGTGCTGCTGGTCTGGAACAGTGGGTTTCATC 1844
Db 1875 GAATGGCTGGGCTCTTACTGCGCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1934
Qy 1845 GTGTCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1904

Db 1935 CTGCACCTCTTGTCTGCTGCTTACTATATATTGACGAGATTTACAGAACTTCGACCTCCCTG 1994
Qy 1905 TCCACCTGACACCTACCTGCTCCATACATACAGTCTATGGCAAGAGCTTGTATTCATG 1964
Db 1995 CCCCCCTAAACACAAATCTGAAAGCCACAGCTTATGGTGTCCAGGCTGTGCTGCCCTG 2054
Qy 1965 CGGCTCTGGAGTAAACCAATCAGGACCATTCGGTTTGTATAGTACTGCTTTTCTA 2024
Db 2055 TGTCTCAGGAGCAAGAACCAAGATCCACTCTCTGTGTACAAATGATTTGCACTTCTC 2114
Qy 2025 CCATGAAAAAGAAATCAGATTTTGCACATGACTTTAGCAACCTCAGCAGTGTGGGCTC 2084
Db 2115 ACGCAACACTCCCAACAGGACTTTCAACTACAACTTCTCCGCTTGGCAACACCGCTCAC 2174
Qy 2085 ATTAATGAATGGCCCCAGCTTCACTCCAAAGGAACAAATACCTTCCATTTCTTCAATAT 2144
Db 2175 TCTTGTCTGGAGGCCAGCTTCACTTCCAAAGGTTGAAATACATTTCCATCACTTTACCT 2234
Qy 2145 CAGTTTATGTGGGATGAGGGAAGAAGTGGCTCTCTGTACCAACAATATATACAGACTT 2204
Db 2235 CAGTCTCTGTGAAACAGGAGGTAGGAAATGTCTGTGTGCAACGACAAATGTCACTGACCT 2294
Qy 2205 TACAGTAAAGAAATAGTGGCAGGCTCAGATGATTACACAAATTTGGTAGGGGCAATTTG 2264
Db 2295 CCGGATTCCTGAGGTGAGTCAGGG-----TCTCCAAATCTATCACAGCTCAGT 2345
Qy 2265 ATGCCAGTCAACAATATTATTCCTTCTGAAAGTAAGGTTTCCGAGCAGCTTTATCATCACA 2324
Db 2346 CTGCCAGCAGTATCATCTCCCCAGAGGTGACAGGCTTACAAGCGCGGGTTTCTCTACA 2405
Qy 2325 ATCCATCATCTTGGCAGATCATATCAGAGTACAGCTTGAACACCATTTGAAANAATAT 2384
Db 2406 GCCTGTGACGCTTGTGATGACTTATTTGGGTGACAAAGATATGACTCTGGATGGAAT 2465
Qy 2385 TAATATAAAGAGATATGTTCCCACTGTTCCCAAGCAACATACCAAGATGTGCATTTCTT 2444
Db 2466 CACCTCCCGAGTGAACCTTTTCCACCTGGAGTCTTGGAAATACCGAGCTGATCTTCT 2525
Qy 2445 TTATAAGTCTTCTACAGCAACAACATCTTGTATTAATGCCCAGATCACTGCTGTGAAAT 2504
Db 2526 TTATAGTCCAATGATGTGACCCAGTCTCTGCACTTCTGGAGATCAACACCATCCGGT 2585
Qy 2505 GAGGTGTAATCTACTAATCTGGAGCAGGAGTGTTCAGTCCCGCCAGCAAGTGCACG 2564
Db 2586 CAGGTGCACTGCACAGAAACCTGCTCCCTGGAAGTTGCTGTGCCAGAACGCTGTCTAGA 2645
Qy 2565 AGTACTCTGATGGGTGTACCTTCTATTTCTGTGGAGAGTGTGAAAGTGTGCTTCCCT 2624
Db 2646 TGGAGCTGTGATGCTGCACTTCCACTTCTGTGGAGAGCGGCTGTGCTGCCGCT 2705
Qy 2625 GTGTACGGAGCATGATCTCCATGAGATTCAGGAGGCTTCCAGAGAGGATTTTCAGAAAC 2684
Db 2706 CTGCTCAGTGGCTGACTTACCATGCTATCTGACAGCTGTGTGGCTGGGATCCAGANGAC 2765
Qy 2685 CTTGTATGTGGAATGAACCTTAATGTTGTCATTAAGGAATTTCTTCCCTGAGAAAAA 2744
Db 2766 TACTTACGTGTGNCGAGAACCCCAAGTATGCTCTGGTGGCAATTTCTGCTGTGAGCAGAG 2825
Qy 2745 GTTGGCAACTGTGAAACGTTTCACTTTTGGCTGAAAGTGGAGCGGCTGTGGAGCTTT 2804
Db 2826 AGTCACCATCTGCAAAACCATAGATTTCTGCTGAAAGTGGGATCTCTGCAAGGACCTG 2885
Qy 2805 TACTGCGGTTTGGCTGGCTGTGACTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2864
Db 2886 TACTGCCATCTGCTCAGCGCTTGTGACCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2945
Qy 2865 AFACAATATTCAAGTTAGTATGAGCTACTCAAAAGAGTGTGAATCTCCCGCTGC 2924
Db 2946 GTACAAGTACTTCAAGCTGGTGTGATGATGCTCTCTCAAGAGCTGTGACTGCCAGCAGC 3005
Qy 2925 AGACAGTTGTCTCATGGAAGGAGAAATATTAAGAGGAAAGTTGTATTTTCCAAATA 2984
Db 3006 TGACAGCTGCGCCATCATGGAAGCGGAGGATGTAGAGGAGCACTCATCTTTTACCAGCA 3065

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 00:01:37 ; Search time 8400.33 Seconds
(without alignments)
11744.602 Million cell updates/sec

Title: US-10-073-333A-1
Perfect score: 3390
Sequence: 1 atgtgttcctccgcgcgggg.....ggtataaaaaaaaaaaaaa 3390

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg_hum.*
31: em.htg_inv.*
32: em.htg_other.*
33: em.htg_mus.*
34: em.htg_pln.*
35: em.htg_rtd.*
36: em.htg_mam.*
37: em.htg_vrt.*
38: em_sy.*
39: em.htgo_hum.*
40: em.htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2617.2	77.2	3185	9	AK055902	AK055902 Homo sapi
2	1499.6	44.2	1737	6	AX060311	AX060311 Sequence
3	1182.6	34.9	1508	6	AX060301	AX060301 Sequence
4	904.6	26.7	3280	6	AX036667	AX036667 Sequence
5	903	26.6	3331	6	AX127726	AX127726 Sequence
6	901	26.6	3501	6	AX463904	AX463904 Sequence
7	860.8	25.4	2610	6	AX127742	AX127742 Sequence
8	860.8	25.4	2733	6	AX127740	AX127740 Sequence
9	709.6	20.9	2895	9	AK057647	AK057647 Homo sapi
10	600.2	17.7	2449	9	BC031648	BC031648 Homo sapi
11	547.	16.1	2129	9	AK026832	AK026832 Homo sapi
c 12	517.6	15.3	136150	9	AC002081	AC002081 Homo sapi
13	517.2	15.3	5567	9	AB037745	AB037745 Homo sapi
14	483.2	14.5	1129	6	AR184096	AR184096 Sequence
c 15	482.2	14.2	576	6	AX336856	AX336856 Sequence
16	400.4	11.8	2188	10	BC029010	BC029010 Mus muscu
17	400.2	11.8	1717	6	AX013071	AX013071 Sequence
18	381.8	11.3	1149	6	AX127746	AX127746 Sequence
19	285.8	8.4	1119	6	AX213279	AX213279 Sequence
20	285.8	8.4	1587	6	AX213277	AX213277 Sequence
c 21	265.4	6.9	627	6	AX127744	AX127744 Sequence
22	234.6	6.4	131321	2	AC080160	AC080160 Mus muscu
23	217	5.8	404	6	AX071721	AX071721 Sequence
24	198	5.8	101719	2	AL357495	AL357495 Homo sapi
25	196.4	5.8	1149	10	BC022655	BC022655 Mus muscu
26	196.2	5.7	165865	2	AC079958	AC079958 Mus muscu
c 27	192.8	5.6	107889	9	AC004519	AC004519 Homo sapi
c 28	189	5.3	108063	2	AC116297	AC116297 Rattus no
c 29	180.4	4.8	426	6	AX368203	AX368203 Sequence
c 30	161.6	4.1	437	6	AX182039	AX182039 Sequence
c 31	140.6	4.0	480	6	AX182004	AX182004 Sequence
c 32	135.4	3.8	29979	9	AC000122	AC000122 Homo sapi
c 33	130.4	3.8	131321	2	AC080160	AC080160 Mus muscu
c 34	127.2	3.6	353	6	AX182022	AX182022 Sequence
c 35	123.6	3.5	166384	2	AC113446	AC113446 Mus muscu
c 36	118.6	3.5	108063	2	AC116297	AC116297 Rattus no
c 37	118	3.4	399	6	AX182000	AX182000 Sequence
c 38	115	3.4	183815	2	AC099081	AC099081 Rattus no
c 39	114	3.3	150129	2	AC119789	AC119789 Rattus no
c 40	112.6	3.1	443	6	AX127734	AX127734 Sequence
c 41	104	2.9	97114	9	AL356389	AL356389 Human DNA
c 42	99.8	2.8	677	6	AX036669	AX036669 Sequence
c 43	94.4	2.8	265	6	AX368154	AX368154 Sequence
c 44	94	2.7	89129	10	AL672200	AL672200 Mouse DNA
c 45	90.8					

ALIGNMENTS

RESULT 1

AK055902

LOCUS

DEFINITION

AK055902

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

AK055902 3185 bp mRNA linear PRI 01-AUG-2002
Homo sapiens cDNA FLJ31340 fis, clone MESAN1000035, weakly similar
to MAJOR SURFACE-LABELLED TROPHOZOITE ANTIGEN PRECURSOR.

AK055902.1 GI:16550745
oligo capping: fis (full insert sequence).
Homo sapiens normal mesangial cells (NHMC56046-2) cDNA to mRNA,
clone_lib:MESAN1 clone:MESAN1000035.

Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Tashiro,H., Yamazaki,M., Watanabe,K., Kumagai,A., Itakura,S.,
1

Fukumizu, Y., Fujimori, Y., Komiya, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Waga, T., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahari, K., Masuho, Y., Nagai, K. and Isogai, T.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 3185)
Isogai, T., Otsuki, T. and Sugiyama, T.
Direct Submission
Submitted (24-Oct-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.

FEATURES

Location/Qualifiers

1..3185
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MESAN100035"
/cell_type="normal mesangial cells (NHMC56046-2)"
/clone_lib="MESAN1"
/note="cloning vector: pME18SFL3-primary culture, normal mesangial cells"
362..1843
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB71041.1"
/db_xref="GI:16550746"
/translation="MLEGNQVLMKSGTNILYWRITGILMSKAVKPVLVNITIEGV
AVTSRCFCKPQTFNKGKNCQVCPNTYSEKAKECIRCKDDSPSEGSSECTE
RPTCTIDFQIHTPCDEEGTKIMYKIEWPKICREDLTDAILRPPSGEKDKCPNP
GYNNGSSSCPPGPTSDGTEKPCPAGTEPALGFYKWNVLPGNMKTSCFNVG
NSKCDMGWEVAGHIQSGAGSDNYLILNLHPKFPKPTSMGTGATGSELGRITFV
FETLGSADCLVFMVDINRKNSTNVVSMGTTKEOAYTHIFKNATFTWAFORTNQ
GODNRFFNDMWKYSITATNAVDCVASSCRALGSESSSCVPCPPGHIKETN
QCKECPDPTLSIHQVYKKEACIPCGPSKQNDHSVCYSDCFIPEKENSLSHYDFS
NLSVSGSLMNGSPFSKGTFFHFNISIVGMRRRLVSVPTI"

CDS

BASE COUNT
ORIGIN

960 a 660 c 719 g 846 t

Query Match 77.2%; Score 2617.2; DB 9; Length 3185;
Best Local Similarity 93.2%; Pred. No. 0;
Matches 2866; Conservative 0; Mismatches 3; Indels 207; Gaps 4;

QY 504 CAACCTCTTGTGATCCCTCGTGGAACTACATAGAACTAAATCGTGATGACTGCACCGGT 563
DB 111 CAGCTCTTGTGATCCCTCGTGGAACTACATAGAACTAAATCGTGATGACTGCACCGGT 170
QY 564 GTCTTTGATCTATGCTGTCACCTTAAGAAGTCAGGCTATGCTTCTTTGAGTACACAGTA 623
DB 171 GTCTTTGATCTATGCTGTCACCTTAAGAAGTCAGGCTATGCTTCTTTGAGTACACAGTA 230
QY 624 TGTGCAACAACAACATCTTCTTTGAGTCTTTTATTCAAAATGATCATGTCAGGAGATGGA 683
DB 231 TGTGCAACAACAACATCTTCTTTGAGTCTTTTATTCAAAATGATCATGTCAGGAGATGGA 290
QY 684 CACCACCACTGACAGTGGGTAAAACCTTACAGACAAATGGAGAAATGGGGCTCTCATTTCT-- 741
DB 291 CACCACCACTGACAGTGGGTAAAACCTTACAGACAAATGGAGAAATGGGGCTCTCATTTCTGA 350
QY 742 -----GTAATCCTGAAATCAGGCACAACATAC 769
DB 351 CAGCTCCTTAATGCTTGAAGCAATTAACAGGTAACTGCTGAAATCAGGCACAACATAC 410

QY 770 TCTACTGGAGAACTACAGGCATCCTTATGGGTCTTAAGCGGTCAAGCCTGCTGTGTTAA 829
DB 411 TCTACTGGAGAACTACAGGCATCCTTATGGGTCTTAAGCGGTCAAGCCTGCTGTGTTAA 470
QY 830 AAAATATCAAAATTGAAGGGGTGGCGTACACATACAAATGTTTCTTCCAGCAGCCAGCA 889
DB 471 AAATATCAAAATTGAAGGGGTGGCGTACACATACAAATGTTTCTTCCAGCAGCCAGCA 530
QY 890 CATTCAGCAACAACCAAGGTTTCAATTCAGCTGCCAGGTGTGTCCAGAAACACCTATTCTG 949
DB 531 CATTCAGCAACAACCAAGGTTTCAATTCAGCTGCCAGGTGTGTCCAGAAACACCTATTCTG 590
QY 950 AGAAGAGCCCAAGAAATGATTAAGGTGTAAGAGCAGCTCTCAATTTTC-----AGGAT 1003
DB 591 AGAAGAGCCCAAGAAATGATTAAGGTGTAAGAGCAGCTCTCAATTTTC-----AGGAT 650
QY 1004 CCAGTCAGTGTACAGAGCCCTCCCTGTACACAAAAGACTATTTCCAGATCCATACAC 1063
DB 651 CCAGTCAGTGTACAGAGCCCTCCCTGTACACAAAAGACTATTTCCAGATCCATACAC 710
QY 1064 CATGTGATGAAGAAAGAGACACAGATTAATGTACAAGTGGATAGAGCCCAAAATCTGCC 1123
DB 711 CATGTGATGAAGAAAGAGACACAGATTAATGTACAAGTGGATAGAGCCCAAAATCTGCC 770
QY 1124 GGGAGGATCTACAGATGCTATTAGATTGCCCTTCTGGAGAGAAGAGGATTGTCGCG 1183
DB 771 GGGAGGATCTACAGATGCTATTAGATTGCCCTTCTGGAGAGAAGAGGATTGTCGCG 830
QY 1184 CTTGCAACCTCGATTATTAACAATGGATCATCTTCTGGCATCCCTGTCTCTCGGAA 1243
DB 831 CTTGCAACCTCGATTATTAACAATGGATCATCTTCTGGCATCCCTGTCTCTCGGAA 890
QY 1244 CATTTTCAGATGGAACCAAGAAATGTAGACCATGTCCAGCAGGAGCGGCTGCACATTG 1303
DB 891 CATTTTCAGATGGAACCAAGAAATGTAGACCATGTCCAGCAGGAGCGGCTGCACATTG 950
QY 1304 GCTTTGAATATAAATGGTGAATGTCTTCTGGCAACATGAAATCTTCTGCTTCAATG 1363
DB 951 GCTTTGAATATAAATGGTGAATGTCTTCTGGCAACATGAAATCTTCTGCTTCAATG 1010
QY 1364 TTGGCAATCAAGTGCATGGAATGATGGTGGAGAGTGGCTGGAGATCATATCCAGA 1423
DB 1011 TTGGCAATCAAGTGCATGGAATGATGGTGGAGAGTGGCTGGAGATCATATCCAGA 1070
QY 1424 GTGGGCTGGAGTGTCTGACAAATGATTAACCTTAACTTGATATCCACAGATTTA 1483
DB 1071 GTGGGCTGGAGTGTCTGACAAATGATTAACCTTAACTTGATATCCACAGATTTA 1130
QY 1484 AACCAACAATCTATGACTGGAGCCAGGGTCTTGAACCTAGGAAGATAACATTTGCT 1543
DB 1131 AACCAACAATCTATGACTGGAGCCAGGGTCTTGAACCTAGGAAGATAACATTTGCT 1190
QY 1544 TTGAGACCTCTGTTCAGCTGCTGCTGTTTGTACTTCTATCATGGTGGATTAATAAGAAAA 1603
DB 1191 TTGAGACCTCTGTTCAGCTGCTGCTGTTTGTACTTCTATCATGGTGGATTAATAAGAAAA 1250
QY 1604 GTACAAATGTGGTGAATCGTGGGTGGACCAAGAAAAACAAGCTTACACCCATATCA 1663
DB 1251 GTACAAATGTGGTGAATCGTGGGTGGACCAAGAAAAACAAGCTTACACCCATATCA 1310
QY 1664 TCTTCAAGAATGCAACTTTTACATTTACATGGGGATTTCCAGAGAACTTAATCAGGGTCAAG 1723
DB 1311 TCTTCAAGAATGCAACTTTTACATTTACATGGGGATTTCCAGAGAACTTAATCAGGGTCAAG 1370
QY 1724 ATAATAGACGGTTCATCAATGACATGTTGAAGATTTATCTATCACAGCCACTAATCCAG 1783
DB 1371 ATAATAGACGGTTCATCAATGACATGTTGAAGATTTATCTATCACAGCCACTAATCCAG 1430
QY 1784 TTGATGGGGTGGCGTCTTCATGCGCGTCCCTGTGCGCTTCTTGAACAGTCCGGTTCAT 1843
DB 1431 TTGATGGGGTGGCGTCTTCATGCGCGTCCCTGTGCGCTTCTTGAACAGTCCGGTTCAT 1490
QY 1844 CGTGTGTCCTGCCCTCCAGGCCACTACATTGAGAAAGAAACCAACCAAGTGCAGAGGAT 1903

```

|||||
Db 1491 CGTGTGTCCTCCCTCCAGGCCACTACATTGAGAAAGAAACCAACCAAGTCAAGGAAT 1550
QY 1904 GTCCACCTCACACCTACCTGTCCATACATCAGGTCTATGCGCAAGAGGCTTGATTCAT 1963
Db 1551 GTCCACCTCACACCTACCTGTCCATACATCAGGTCTATGCGCAAGAGGCTTGATTCAT 1610
QY 1964 GCGGCGCTGGGAGTAAACAATCAGGACCATTCGGTTTGCCTATAGTACTGCTTTTCT 2023
Db 1611 GCGGCGCTGGGAGTAAACAATCAGGACCATTCGGTTTGCCTATAGTACTGCTTTTCT 1670
QY 2024 ACCATGAAAAGAAATCAGATTTTGCACCTATGACTTTAGCAACCTCAGCAGTGGGGT 2083
Db 1671 ACCCTGAAAAGAAATCAGATTTTGCACCTATGACTTTAGCAACCTCAGCAGTGGGGT 1730
QY 2084 CATTAAATGAATGGCCCGCCAGCTTCACCTCCAAAGGAACAAAATACCTCCATTCTTCAATA 2143
Db 1731 CATTAAATGAATGGCCCGCCAGCTTCACCTCCAAAGGAACAAAATACCTCCATTCTTCAATA 1790
QY 2144 TCAGTTTATGTGGGCATGAGGGGAAGAAGATGGCTCTCTGTACCAACATATAACAGACT 2203
Db 1791 TCAG-TTATGTGGGCATGAGGGGAAGAAGATGGCTCTCTGTACCAACATATAACAGACT 1849
QY 2204 TTACAGTAAAGAAATAGTGGCAGGGTCAGATGATTACACAAATTTGGTAGGGGATTTG 2263
Db 1850 TTACAGTAAAGAAATAGTGGCAGGGTCAGATGATTACACAAATTTGGTAGGGGATTTG 1909
QY 2264 TATGCCAGTCAACAATATTCTCTCTGAAAGTAAGGTTTCCGACGAGCCTTATCATCAC 2323
Db 1910 TATGCCAGTCAACAATATTCTCTCTGAAAGTAAGGTTTCCGACGAGCCTTATCATCAC 1969
QY 2324 AATCCATCAATCTGCAGATACATTCATAGGAGTCACAGTTGAAACCCACATTCGAAAAATA 2383
Db 1970 AATCCATCAATCTGCAGATACATTCATAGGAGTCACAGTTGAAACCCACATTCGAAAAATA 2029
QY 2384 TTAATATAAAGAAATATGTTCCAGTTCACCAAGCCAAATACAGATGTCATTTCT 2443
Db 2030 TTAATATAAAGAAATATGTTCCAGTTCACCAAGCCAAATACAGATGTCATTTCT 2089
QY 2444 TTATAAGTCTTCTACAGCAACAACATCTTGTATTAAATGGCCGATCACTGCTGTGAAAA 2503
Db 2090 TTATAAGTCTTCTACAGCAACAACATCTTGTATTAAATGGCCGATCACTGCTGTGAAAA 2149
QY 2504 TGAGTGTAAATCCTACTAAATCTGGAGCAGGAGTGATTTAGTCCCCAGCAAGTGCCCGAG 2563
Db 2150 TGAGTGTAAATCCTACTAAATCTGGAGCAGGAGTGATTTAGTCCCCAGCAAGTGCCCGAG 2209
QY 2564 CAGGTACCTGTGATGGGTGATGTTCTATTTCTGTGGGAGAGTGTGAAGCTTGCCCTC 2623
Db 2210 CAGGTACCTGTGATGGGTGATGTTCTATTTCTGTGGGAGAGTGTGAAGCTTGCCCTC 2269
QY 2624 TGTGTACGGAGCATGACTTCCATGAGATTGAGGAGCCCTGCAAGAGAGGATTTTCAGGAAA 2683
Db 2270 TGTGTACGGAGCATGACTTCCATGAGATTGAGGAGCCCTGCAAGAGAGGATTTTCAGGAAA 2329
QY 2684 CTTGTATGTGTGGAATGAACCTAAATGGTGCAATTAAGGAATTTCTTTGCTGAGAAAA 2743
Db 2330 CTTGTATGTGTGGAATGAACCTAAATGGTGCAATTAAGGAATTTCTTTGCTGAGAAAA 2389
QY 2744 AGTTGGCAACCTGTGAACGGTTGACTTTTGGCTGAAGTGGGAGCCGCTGGGAGCTT 2803
Db 2390 AGTTGGCAACCTGTGAACGGTTGACTTTTGGCTGAAGTGGGAGCCGCTGGGAGCTT 2449
QY 2804 TTACTGCCGCTTTTGTGGTGGCTCTGACCTGCTACTTCTGAAAAAGAAATCAAAA- 2858
Db 2450 TTACTGCCGCTTTTGTGGTGGCTCTGACCTGCTACTTCTGAAAAAGAAATCAAAA 2509
QY 2859 ----- 2858
Db 2510 AATACAAATATTCCAAGTTAGTAAATGACGACTAACTCAAAAGAGTGTGAATCCCGGCTG 2569
QY 2859 ----- 2858
```

```

Db 2570 CAGACAGTTGTGTCTATCATGGAAGGAGAAGATAATCAGAGGAAGTTGTATATTCCAAATA 2629
QY 2859 -----GAAAAGAGACCATTTTG 2877
Db 2630 ACAGTCACTACTAGGAAAACCTCAAATCTTGGCAACCAAGGAAAAGAGACCATTTTG 2689
QY 2878 AATCTGTTCAACTGAAAACCTCAAAGATCCCAAAATATATGAAGAGACAGTGTGTAGCCT 2937
Db 2690 AATCTGTTCAACTGAAAACCTCAAAGATCCCAAAATATATGAAGAGACAGTGTGTAGCCT 2749
QY 2938 TGAGACTAATGAACAAAGAAACCTGCTCTAGTTTTCAGAGACCATATTTTAGGGTCTGTC 2997
Db 2750 TGAGACTAATGAACAAAGAAACCTGCTCTAGTTTTCAGAGACCATATTTTAGGGTCTGTC 2809
QY 2998 CTCATACCTGTCACATTTGCTGATCTCACAGAGGAGGCCATGCCGTGAAAAGGAAGA 3057
Db 2810 CTCATACCTGTCACATTTGCTGATCTCACAGAGGAGGCCATGCCGTGAAAAGGAAGA 2869
QY 3058 GATTGAAAACATTTGCTTATCACATGGTCAAGTACCTTGCCTTGTCCAAATAAAGGAAAAGCA 3117
Db 2870 GATTGAAAACATTTGCTTATCACATGGTCAAGTACCTTGCCTTGTCCAAATAAAGGAAAAGCA 2929
QY 3118 AATGATTTGGGTCTCAACTGAAGATGAAGTCAACTCAGAGAGAGATTTATCTGTATATA 3177
Db 2930 AATGATTTGGGTCTCAACTGAAGATGAAGTCAACTCAGAGAGAGATTTATCTGTATATA 2989
QY 3178 CACATAACTGAAAACCAAGTTTAAAGCCCAACCAATGCACCTGCTGATGCATGCCATATAAT 3237
Db 2990 CACATAACTGAAAACCAAGTTTAAAGCCCAACCAATGCACCTGCTGATGCATGCCATATAAT 3049
QY 3238 AATGGTAACCTTTTATCTTTATGATGCTACATAACAAGTGTGATTTGGAAGGCACATG 3297
Db 3050 AATGGTAACCTTTTATCTTTATGATGCTACATAACAAGTGTGATTTGGAAGGCACATG 3109
QY 3298 TGAGCATATGCAATATGATCCAAATTTATGTTTCTTTGTTTATATTTTGGGAAAAAT 3357
Db 3110 TGAGCATATGCAATATGATCCAAATTTATGTTTCTTTGTTTATATTTTGGGAAAAAT 3169
QY 3358 AAAATTTTTTTAAAGT 3373
Db 3170 AAAATTTTTTTAAAGT 3185
```

```

RESULT 2
AX060311
LOCUS AX060311 1737 bp DNA linear PAT 22-JAN-2001
DEFINITION Sequence 19 from Patent WO0078802.
ACCESSION AX060311
VERSION AX060311.1 GI:12405800
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1737)
AUTHORS Shimkets,R.A., Fernandes,E., Vernet,C., Yang,M., Boldog,F.L. and
Herrmann,J.L.
TITLE Secreted polypeptides and corresponding polynucleotides
JOURNAL Patent: WO 0078802-A 19 28-DEC-2000;
Curagen Corporation (US)
FEATURES
Location/Qualifiers
source 1..1737
/organism="Homo sapiens"
/db_xref="taxon:9606"
296..1690
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC24875.1"
/db_xref="GI:12405801"
/translation="MKNQVCSKCGEYSLGSGIKFDEWDELPAFNSIATFMDTVVG
PDSRPDCNNSWIPRNYIESNDDCTVSLIIYAVHLKSGYVFEFYQYVNNLIFFE
FFIQNQCQEMDTTDDKVLKLTNGENSGSHVYMLKSGNTILYWRITGLMDSKVKPY
LVKNITIEGVAYTSECFPCPKPTFSNKPQSFNCQVCPRNTYSEKAKECIRCKDQSQF
```

SEBSSBCTRRPCTTKDYFOIHTPCDEBCKTQIMYKWIPEKICREDLTDAILRPPSG
EKDCPPCNPFYNGSSCHPDPGTFSDGTEKRCPCPAGTEPDPALGFYKWNVLPG
NMKTSFNVGSKDGNHWEVAGDHIQSGAGSDNDYLLNLHPGFKPPTSMGTGAT
GSELGRITFFVFETLCSADCVLYEMVDINRKSNTNVESWGKTKEQOAYTHIIFKNATET
FTWGIPELIIQGPB"
BASE COUNT 495 a 381 c 409 g 451 t 1 others
ORIGIN

Query Match 44.2%; Score 1499.6; DB 6; Length 1737;
Best Local Similarity 98.7%; Pred. No. 0;
Matches 1545; Conservative 0; Mismatches 10; Indels 10; Gaps 3;
QY 189 GAAAGATTATCACTTTGAATATACCGAATGATAGCAGTGGCTCCAGTGAGAGTTGCC 248
DB 142 GAAAGATTATCACTTTGAATATACCGAATGATAGCAGTGGCTCCAGTGAGAGTTGCC 201
QY 249 CATTCCAAATCTCGAGTGGAGCTGCTGGCGCTGCTGACCCAGTGAGAGGCAAGAATG 308
DB 202 CATTCCAAATCTCGAGTGGAGCTGCTGGCGCTGCTGACCCAGTGAGAGGCAAGAATG 261
QY 309 CACTTTCTCTGCTTCTGGAGAGTATCTAGAAATGAAGAACCCAGGTATGCGAATGTCG 368
DB 262 CACTTTCTCTGCTTCTGGAGAGTATCTAGAAATGAAGAACCCAGGTATGCGAATGTCG 321
QY 369 TGGTGAAGGCACCTATCTCTGGGAGTGGCATCAATTTGATGAATGGGATGAATGGC 428
DB 322 TGGTGAAGGCACCTATCTCTGGGAGTGGCATCAATTTGATGAATGGGATGAATGGC 381
QY 429 GCGAGATTTTCTAACATCGCAACATTCATGACACTGTGGTGGCGCTTCTGACAGCAG 488
DB 382 GCGAGATTTTCTAACATCGCAACATTCATGACACTGTGGTGGCGCTTCTGACAGCAG 441
QY 489 GCGAGAGGCTGTACAACTCTTCTGGATCCCTCGTGGAACTACATAGAACTAAATGG 548
DB 442 GCGAGAGGCTGTACAACTCTTCTGGATCCCTCGTGGAACTACATAGAACTAAATGG 501
QY 549 TGATGACTGCAGGTGCTTTGATCTATGCTGTGACCTTAAGAACTAGGCTATGCTCT 608
DB 502 TGATGACTGCAGGTGCTTTGATCTATGCTGTGACCTTAAGAACTAGGCTATGCTCT 561
QY 609 CTTTGAGTACCACTATGTCGACAACAACATCTCTTTGAGTCTTTATTCAAAATGATCA 668
DB 562 CTTTGAGTACCACTATGTCGACAACAACATCTCTTTGAGTCTTTATTCAAAATGATCA 621
QY 669 GTGCCAGGAGATGGACACCACCTGACAAGTGGGTAAACTTACAGACAATGGAGATG 728
DB 622 GTGCCAGGAGATGGACACCACCTGACAAGTGGGTAAACTTACAGACAATGGAGATG 681
QY 729 GGGCTCTCATTCTGTAATGCTGAATCAGGCACAACATCTCTACTGGAGAACTACAGG 788
DB 682 GGGCTCTCATTCTGTAATGCTGAATCAGGCACAACATCTCTACTGGAGAACTACAGG 741
QY 789 CATCTCTATGGGTCTTAAGCGGTCAAGCCTGTGCTGTTGTTGTTTATTCAAAATGATCA 848
DB 742 CATCTCTATGGGTCTTAAGCGGTCAAGCCTGTGCTGTTGTTTATTCAAAATGATCA 801
QY 849 GTGGCGGTACACATCAGATGTTTCTTTCGCAAGCCAGCAGCATTCAGCAACAACCCAGG 908
DB 802 GTGGCGGTACACATCAGATGTTTCTTTCGCAAGCCAGCAGCATTCAGCAACAACCCAGG 861
QY 909 TTCATTCAACTGCCAGGTGCTCCCAAGAACACCTTATCTGCAAGAGGAGCCAAAGAATG 968
DB 862 TTCATTCAACTGCCAGGTGCTCCCAAGAACACCTTATCTGCAAGAGGAGCCAAAGAATG 921
QY 969 TATAAGGTGAAGAGCAGCTCTCAATTTTC-----AGGATCCAGTGAAGTACAGAGCG 1022
DB 922 TATAAGGTGAAGAGCAGCTCTCAATTTTCAGAGGAAGATCCAGTGAAGTACAGAGCG 981
QY 1023 CCTCTCCTGTACCAAAAAGACTATTTCCAGATCCATCTCCTGATGATGAGAGGAA 1082
DB 982 CCTCTCCTGTACCAAAAAGACTATTTCCAGATCCATCTCCTGATGATGAGAGGAA 1041
QY 1083 GACACAGATAATGTACAAGTGGATAGAGCCCAAAAATCTCCCGGGAGGATCTCACAGATGC 1142

DB 1042 GACACAGATAATGTACAAGTGGATAGAGCCCAAAAATCTCCCGGGAGGATCTCACAGATGC 1101
QY 1143 TATTAGATTGCCCCCTTCTGGAGAGAAGAAGATTGTCCGCCCTTCAACCCCTGGATTGTA 1202
DB 1102 TATTAGATTGCCCCCTTCTGGAGAGAAGAAGATTGTCCGCCCTTCAACCCCTGGATTGTA 1161
QY 1203 TAACAAATGATCATCTTCTGCCATCCCTGCTGCTCTGGAACATTTTCAGATGGAAACCA 1262
DB 1162 TAACAAATGATCATCTTCTGCCATCCCTGCTGCTCTGGAACATTTTCAGATGGAAACCA 1221
QY 1263 AGAATGTAGACCATCTCCAGCAGGAAGAGGCTGCACCTTGGCTTTGAATATAAATGGTG 1322
DB 1222 AGAATGTAGACCATCTCCAGCAGGAAGAGGCTGCACCTTGGCTTTGAATATAAATGGTG 1281
QY 1323 GAATGCTCTTCTGCGCAACATGAAAACCTTCCCTGCTTCAATGTTGGGAATTCAAAAGTGCA 1382
DB 1282 GAATGCTCTTCTGCGCAACATGAAAACCTTCCCTGCTTCAATGTTGGGAATTCAAAAGTGCA 1341
QY 1383 TGGATGAATGTTGGGAGTGGCTGGAGATCATATCCAGAGTGGGCTGGAGGTTCTGCA 1442
DB 1342 TGGATGAATGTTGGGAGTGGCTGGAGATCATATCCAGAGTGGGCTGGAGGTTCTGCA 1401
QY 1443 CAATGATTACCTGATCTTAAACTTGCATATCCAGAGTAAACCCACCACCAATCTATGAC 1502
DB 1402 CAATGATTACCTGATCTTAAACTTGCATATCCAGAGTAAACCCACCACCAATCTATGAC 1461
QY 1503 TGGAGCCAGGGTCTGAACTAGGAAGATAACATTTGCTTTTGAGACCCCTCTGTTTCAGC 1562
DB 1462 TGGAGCCAGGGTCTGAACTAGGAAGATAACATTTGCTTTTGAGACCCCTCTGTTTCAGC 1521
QY 1563 TGACTGTGTTTTGCTACTTCATGCTGGATATTAATAGAAAAAGTACAAAATGCTGGTAGAATC 1622
DB 1522 TGACTGTGTTTTGCTACTTCATGCTGGATATTAATAGAAAAAGTACAAAATGCTGGTAGAATC 1581
QY 1623 GTGGGTGGAACCAAGAAAAACAAGCTTACACCCATATCATCTTCAAGAAATGCAACTTT 1682
DB 1582 GTGGGTGGAACCAAGAAAAACAAGCTTACACCCATATCATCTTCAAGAAATGCAACTTT 1641
QY 1683 TACATTTACATGGG--CATTCAGAGAACTAA--CAGGGTCAAGATAATAGACGGTTCA 1738
DB 1642 TACATTTACATGGGCAATTCACAGAGAACTAATTCAGGGTCCCAAGATAATAGACGGTTCC 1701
QY 1739 TCAAT 1743
DB 1702 NCCAT 1706
RESULT 3
AX060301
LOCUS AX060301 1508 bp DNA linear PAT 22-JAN-2001
DEFINITION Sequence 9 from Patent WO0078802.
ACCESSION AX060301
VERSION AX060301.1 GI:12405790
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1508)
AUTHORS Shimkets, R.A., Fernandes, E., Vernet, C., Yang, M., Boldog, F.L. and
Herrmann, J.L.
TITLE Secreted polypeptides and corresponding polynucleotides
JOURNAL Patent: WO 0078802-A 9 28-DEC-2000;
Curagen Corporation (US)
FEATURES
Location/Qualifiers
1..1508
/organism="Homo sapiens"
/db_xref="taxon:9606"
226..1461
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC24870.1"

```
/db_xref="GI:12405791"
/translation="MHSSWIPRGNVIESNRDDCTVSLIYAVHLKSKSYVFEYQYVDN
NIEPFIQNDQCOEMDTTDDKWLKTGEMGSHVMLKSGTNYLWRTYGLMSK
AVRVLVKNITIEGVATVSCFPCKPGTFSNKGPSNQCVPNTYSEKAGRECKR
DDQFSEGSSECTEPPTTKDYFIHTPCDEEGKTQIMYKWEIPKICREDLDAIR
LPPSEKDKPCPCNPGYNNSSCHPCPGTFSNKGPSNQCVPNTYSEKAGRECKR
NVLPGNMTSCFNVNSKCDGMNGWEVAGDHIQSGAGSDNDYLLMLHPGPKPPTS
MTGATSELGRITFVFETLGSADCVLFVMDINKSTNVVSWGGTKEKQAYTHIFK
NATFTWGIPIRELIOGPR"
BASE COUNT      444 a   306 c   355 g   402 t      1 others
ORIGIN
Query Match      34.9%; Score 1182.6; DB 6; Length 1508;
Best Local Similarity 98.4%; Pred. No. 3.1e-266;
Matches 1228; Conservative 0; Mismatches 10; Indels 10; Gaps 3;
QY 506 ACTCTTTTGGATCCCTCGTGGAACACTACATAGATCTAAATCGTGATGACTGCACGGTGT 565
Dy 230 ACTCTTTTGGATCCCTCGTGGAACACTACATAGATCTAAATCGTGATGACTGCACGGTGT 289
QY 566 CTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGCTATGTCTCTTTCAGTACCAAGTATG 625
Dy 290 CTTTGATCTATGCTGTGCACCTTAAGAAGTCAGGCTATGTCTCTTTCAGTACCAAGTATG 349
QY 626 TCGACAACAACATCTCTTTGAGTTCTTTATTCAAAATGATCAGTCCAGGAGATGGACA 685
Dy 350 TCGACAACAACATCTCTTTGAGTTCTTTATTCAAAATGATCAGTCCAGGAGATGGACA 409
QY 686 CCACACTGACAAGTGGGTAAACCTTACAGACATGGAGATGGGGCTCTCATTTCTGTAA 745
Dy 410 CCACACTGACAAGTGGGTAAACCTTACAGACATGGAGATGGGGCTCTCATTTCTGTAA 469
QY 746 TGCTGAAATCAGGCACAAACATCTCTACTGGAGAACTACAGGCATCTTATGGTGTCTTA 805
Dy 470 TGCTGAAATCAGGCACAAACATCTCTACTGGAGAACTACAGGCATCTTATGGTGTCTTA 529
QY 806 AGCGGTCAAGCTGTGCTGTTGTTAAATAATACAAATGAAAGGGTGGCGTACACATCAG 865
Dy 530 AGCGGTCAAGCTGTGCTGTTGTTAAATAATACAAATGAAAGGGTGGCGTACACATCAG 589
QY 866 AATGTTTTCTTTCGACGACGACATTCAGCAACAACACAGTTTCATTCACCTCCACGG 925
Dy 590 AATGTTTTCTTTCGACGACGACATTCAGCAACAACACAGTTTCATTCACCTCCACGG 649
QY 926 TGTGTCCCAAGAACACCTATTCTGAGAAAGGAGCAAAAGATGTATAGGTGTAAAGACG 985
Dy 650 TGTGTCCCAAGAACACCTATTCTGAGAAAGGAGCAAAAGATGTATAGGTGTAAAGACG 709
QY 986 ACTCTCAATTTTC-----AGGATCCAGTGTGTACAGAGCGCCCTCCCTGTACCACAA 1039
Dy 710 ACTCTCAATTTTCAGAGGAAGGATCCAGTGTGTACAGAGCGCCCTCCCTGTACCACAA 769
QY 1040 AAGACTATTTCCAGATCCATCTCATGTATGAAGAAGGAAGACACAGATAATGTACA 1099
Dy 770 AAGACTATTTCCAGATCCATCTCATGTATGAAGAAGGAAGACACAGATAATGTACA 829
QY 1100 AGTGGATAGAGCCCAAAATCTCGCGGAGGATCTACACATGCTATTAGATTGCCCCCTT 1159
Dy 830 AGTGGATAGAGCCCAAAATCTCGCGGAGGATCTACACATGCTATTAGATTGCCCCCTT 889
QY 1160 CTGGAGAGGAAGGATTTGCGCCCTTGCACCCCTGGATTTTATAACAATGGATCATCTT 1219
Dy 890 CTGGAGAGGAAGGATTTGCGCCCTTGCACCCCTGGATTTTATAACAATGGATCATCTT 949
QY 1220 CTTGCCATCCCTGCTCTCTGAGACATTTTCAGATGGAACCAAGAAATGTACACCATGTC 1279
Dy 950 CTTGCCATCCCTGCTCTCTGAGACATTTTCAGATGGAACCAAGAAATGTACACCATGTC 1009
QY 1280 CAGCAGGAACGGAGCTGACCTTGGCTTTTGAATATATAATGGTGAATGCTCTTCTGGCA 1339
Dy 1010 CAGCAGGAACGGAGCTGACCTTGGCTTTTGAATATATAATGGTGAATGCTCTTCTGGCA 1069
QY 1340 ACATGAAAACCTTCCCTGCTTCAATGTTGGGAATTCAAAGTGGATGGAATGAATGGTGGG 1399
```

```
|||||
Db 1070 ACATGAAAACCTTCTGCTTCAATGTTGGGAATTCAAAAGTCGATGGAATGAATGGTTGGG 1129
QY 1400 AGTGGCTGGAGATCATATCCAGAGTGGGCTGAGGTTCTGACAATGATTTACCTGATCT 1459
Dy 1130 AGTGGCTGGAGATCATATCCAGAGTGGGCTGAGGTTCTGACAATGATTTACCTGATCT 1189
QY 1460 TAAACTTGCATATCCAGGATTTAAACCAACCAACATCTATGACTGGAGCCACGGGTTCTG 1519
Dy 1190 TAAACTTGCATATCCAGGATTTAAACCAACCAACATCTATGACTGGAGCCACGGTTCTG 1249
QY 1520 AACTAGGAAGAATAAATCTTCTTTTGGAGCCCTCTGTTTACAGTGCAGTGTGTTTGTACT 1579
Dy 1250 AACTAGGAAGAATAAATCTTCTTTTGGAGCCCTCTGTTTACAGTGCAGTGTGTTTGTACT 1309
QY 1580 TCATGCTGATATTAATAGAAAAGTACAAATGTTGTTAGAAATCGTGGGTGGACCAAG 1639
Dy 1310 TCATGCTGATTAATAGAAAAGTACAAATGTTGTTAGAAATCGTGGGTGGACCAAG 1369
QY 1640 AAAACAAGCTTACACCCATATCATCTTTCAAGAATGCAACTTTTACATTTACATGGG--C 1697
Dy 1370 AAAACAAGCTTACACCCATATCATCTTTCAAGAATGCAACTTTTACATTTACATGGGCA 1429
QY 1698 ATTCCAGAGAACTAAT--CAGGGTCAAGATAATAGACGGTTCATCAAT 1743
Dy 1430 TTCCAGAGAACTAATTCAGGGTCCAGATAATAGACGGTTCCTCCCAT 1477
RESULT 4
AX036667 AX036667 3280 bp DNA linear PAT 16-NOV-2000
LOCUS Sequence 1 from Patent WO0058460.
DEFINITION AX036667
ACCESSION AX036667
VERSION AX036667.1 GI:11226242
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3280)
AUTHORS Bruck,C.E., Vinals,Y.D., Coche,T. and Cassart,J.P.
TITLE Novel compounds
JOURNAL Patent: WO 0058460-A 1 05-OCT-2000;
BRUCK CLAUDINE ELVIRE MARIE (BE); SMITHKLINE BEECHAM BIOLOG (BE);
VINALS Y D DE BASSOLS CARLOTA (BE); COCHE THIERRY (BE); CASSART
JEAN POL (BE)
FEATURES
source Location/Qualifiers
1..3280
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 810 a 930 c 797 g 743 t
ORIGIN
Query Match 26.7%; Score 904.6; DB 6; Length 3280;
Best Local Similarity 60.1%; Pred. No. 5.4e-201;
Matches 1637; Conservative 0; Mismatches 1059; Indels 27; Gaps 7;
QY 171 GCTTCTCTCTCCAGGAGAAAGATTATCATTGTAATATACGGAATGTGATAGCAGTGG 230
Dy 135 GCTTCTCTCTCCAGGAGAAAGATTATCATTGTAATATACGGAATGTGATAGCAGTGG 194
QY 231 CTCCAGGTGGAGAGTTGCCATTTCCTGAGTGGAGTCTCTGCGCTCCCTGACCC 290
Dy 195 TTCCAGGTGGAGAGTTGCCATTTCCTGAGTGGAGTCTCTGCGCTCCCTGACCC 254
QY 291 AGTGAGAGCAAGAATGCACCTTCTCTGCTCTCTGAGAGATGATCTAGAAATGAAGAA 350
Dy 255 CGTAGGAGCAAGAATGCACCTTCTCTGCTCTCTGAGAGATGATCTAGAAATGAAGAA 314
QY 351 CCAGGTATGCACGTAAGTGTGGTGAAGGACCACTATTCTCTTGGCAGTGGCATCAATTTGA 410
Dy 315 CCAGGTATGCACGTAAGTGTGGTGAAGGACCACTATTCTCTTGGCAGTGGCATTTCCCTCGGACAGGCAATTCGGTTTGA 374
```

QY 411 TGAATGGATGAATTCGCCGACAGGATTTCTAATACGCAACATTCATGGACACTGTGTT 470
Db 375 TGAGTGGATGAGTGCCTCCATGGCTTTGCCAGCCTCTCAGCCAACATGGAGCTGGATGA 434
QY 471 GGGCCCTTCTGACAGGACGACAGGCTGTACAACTCTTCTTTGGATCCCTCGTGGA 530
Db 435 CAGTGTCTGTGAGTGCCA---CCGGGAACCTGTACTTCTCCAAAGTGGTTCCTCCGGGCGCA 491
QY 531 CTACATAGAATCTAATCGTGATGACTGACGGGTGTCTTTGATCTATGCTGTGCACCTTAA 590
Db 492 CTACATCGCTCCCAACAGGACGAATGACAGGCCACACTGATGATGACGCGCTCAACCTGAA 551
QY 591 GAAGTCAAGGTATGTCTTCTTTGAGTACCAATGATGTCGACAAACATCTTCTTTGAGTT 650
Db 552 GCAATCTGGCAGCGCTTAACTTTCGAATACTACTATCCAGACTCCAGCATCATCTTCAGTT 611
QY 651 CTTTATTCAAATGATCAGTGGCAGGAGATGGACACCACTGCAAGTGGGTAAACT 710
Db 612 TTTCTGTTCAATGACAGTGGCAGGCCCAATGCGATGACTC---CAGGTGGATGAAGAC 668
QY 711 TACAGACAATGGAGATGGGGCTCTCATTTCTGTAAATGCTGAATCAGGCACAAACATACT 770
Db 669 CACAGAGAA---AGGATGGGAATTCACAGTGTGGAGCTAAATCGAGGCAATATGTCT 725
QY 771 CTACTGGAGAACTACAGGATCTCTTATGGGTCTTAAGCGGTCAAGCCCTGTGCTGATAA 830
Db 726 CTATTGGAGAACACAGCCTTCTCAGTATGGACCAAAGTACCCAAAGCCTGTGCTGTGAG 785
QY 831 AATATCACAATGGAAGGTGGCTGACATCAGATGATTTTCTGTGCAAGCCAGGAC 890
Db 786 AACAATGGCAATACAGAGGGTGGCTTACACTTCAGAAATGCTTCCCTGCAAACTGGCAC 845
QY 891 ATTCAGCAACAAACAGGTTCATTTCAACTGCCAGTGTGTCAGAAACACCTATTCTGA 950
Db 846 GTATGACAGACAAGAGGGTCTCTTCTGTCAAACTTTGCCAGCCAACTTTATTCAA 905
QY 951 GAAAGGACCAAGATGATAGGTGTAAGTGTAA---GACGACTCTCAATTTTCAGGATCCAG 1007
Db 906 TAAGGAGAACTCTTGGCCACAGTGTGACCTGTGACCTGTGACAAATCTCAGAGAAAGGATCTC 965
QY 1008 TGAGTCTACAGAGCGCCCTCCCTGTACCAACAAGAACTATTTCAGATCCACTCCATG 1067
Db 966 TTCTCTAAGCTGCGGCCAGCTTGCACAGACAAGAAATATTCTACACACACAGCGCTG 1025
QY 1068 TGATGAAGAGGAACACACAGATATGTACAGTGGATAGAGCCCAAAATCTCGCGGA 1127
Db 1026 CGATGCCAACGGAGACACAACACTCATGTACAAATGGGCAAGCCGAAATCTGTAGCA 1085
QY 1128 GGATCTCACAGATGCTTATTTAGATTGGCCCTTCTTGAGAGAGAAGAGGATGTCGCCCTTG 1187
Db 1086 GGACCTTGAGGGGCGAGTGAAGTGCCTGCTGCTGGTGTGAAGACCCACTGCGCCACCTG 1145
QY 1188 CAACCTGGATTTTATACAAATGGATCATCTTCTTGCCATCCCTGTCTCTCGAACATP 1247
Db 1146 CAACCCAGGCTCTTCAAAACCAACAACAGCAGCTGCCAGCCCTGCCATATGGTTCTTA 1205
QY 1248 TTCAGATGCAACCAAGATGTAGACCATGTCCAGCAGGAAGCGGAGCTGCACTTGGCTT 1307
Db 1206 CTCCAATGG---CTCAGACTGTACCGCTGCGCTGCAGGACTGAACCTGTGTGGGATP 1262
QY 1308 TGAATATAAATGGTGAATGTCTTCTTCTGGCAACACTTCCGCTTCAATGTGG 1367
Db 1263 TGAATACAATGGTGAACACGCTGCCCAACAACATGGAACACGCGTCTCAGTGGAT 1322
QY 1368 GAATCAAGTGGATGAATGAATGGTGGAGGTGGCTGGAGATCATATCCAGAGTGG 1427
Db 1323 CAACCTCGAGTACAAGGGCATGACAGGCTGGGAGTGGCTGTGATCATATTCACAGC 1382
QY 1428 GCTGGAGTCTGACATGATTAACCTGATTAACCTGATATCCAGGATTTAAACC 1487
Db 1383 TCGTGGAGCCTCAGACAATGACTTTCATGATTCCTACTCTGTTGTGGCAGGATTTAGACC 1442
QY 1488 ACCAATCTATGACTGGGCCAC---GGTCTCAACTAGGAAGAAATTAACATTTGTCTT 1544

Db 1443 TCCGCACTCGGTGATGGCAGACACAGAGATTAAGAGAGTGGCCAGANTCACATTTGTCTT 1502
QY 1545 TGAGACCCCTCTGTTCAGCTGACTGTGTTTGTACTTTCATGGTGGATTAATATAGAAAAG 1604
Db 1503 TGAGACCCCTCTGTTCGTGAACCTGTGAGCTTACTTTCATGGTGGGTGTAATTCATAGGAC 1562
QY 1605 TACAATGTGGTAGAATCGTGGGTGGAAACCAAGAAAAGAAAGCTTACACCATATCAT 1664
Db 1563 CAACACTCTCTGGAGACGTGAAAGGTTTCCAAAGGCAACAGTCTCTATACCTATCATAT 1622
QY 1665 CTTCAAGAATGCAACTTTTACATTTTACATGGGCATTTCCAGAGAACTTAATCAGGGTCAAGA 1724
Db 1623 TGAGGAGAACACTACCACGAGCTTCACTGGGCTTCCAGAGGACCACTTTTTCATCAGGC 1682
QY 1725 TAATAGACGGTTCATCAATGACATGGTGAAGATTTATTTCTATCACAGGCACATAATGCAAT 1784
Db 1683 AAGCAGGAATGACCAATGAGCTTGGCAAGATCTACTCCATCAATGTCCACCAATGTTAT 1742
QY 1785 TGATGGGTGGGCTCTCATGCGCTGCCCTGCCCTTCTGAACAGTCCGGGTTTCATC 1844
Db 1743 GAATGGGTGGGCTCTACTGCGCTGCCCTCTGCTAGAGCCTCTGATGTGGGCTCTCTC 1802
QY 1845 GTGTGTCCCTGCCCTTCCAGGCCACTTACATTTGAGAAAAGAAACCAACAGTGCAGGAATG 1904
Db 1803 CTGCACCTCTTCTCTGCTGTGTTACTATATTCACCGAGATTCAGGAACCTGCCACTCTCTG 1862
QY 1905 TCCACTGACACTACCTGTCTCATACATCAGTCTATGGCAAGAGGCTGTGATTTCCATG 1964
Db 1863 CCCCCCTAACCAATTTCTGAAAGCCCAACAGCTTATGGTGTCCAGGCTGTGTGCCCTG 1922
QY 1965 CGGGCTGGGAGTAAAAACAATCAGGACCATTCGGTGTGCTATAGTACTGCTGCTTTTCTA 2024
Db 1923 TGGTCCAGGACCAAGACACAAGATCCACTCTCTGTGCTACATGATTCACCTTCTC 1982
QY 2025 CCATGAAAGAAAATCAGATTTTGCATATGACTTTTASCACACTTGCAGAGTGTGGGCTC 2084
Db 1983 AGCAACACTCCCAACAGGACTTCAACTTCTCCGCTTTGGCAACACCGCTCAC 2042
QY 2085 ATTAATGAATGGCCACGCTTCCACTCCAAAGAACAAATATCTTCATTTCTCAATAT 2144
Db 2043 TCTTCTGGAGGCGCAAGCTTCATTTCCAAAGGTTGAATACTTCCATCATCTTTACCT 2102
QY 2145 CAGTTTATGTGGGATGAGGGAAGAAGTGGCTCTCTGTACCAACAATATTAACAGACTT 2204
Db 2103 CAGTCTGTGGAACACAGGAGTAGGAAATGCTGTGTGACCGCAATGTCACTGACT 2162
QY 2205 TACACTAAAGAAATAGTGGCAGGCTCAGATCATTTACACAATTTGGTAGGGCAATTTGT 2264
Db 2163 CCGGATTCCTGAGGTGAGTCAAGG-----TCTCCAAATCTATCACAGCCTACGT 2213
QY 2265 ATGCCAGTCAACAATTTATTTCTTCTGAAAGTANGGTTTCCGAGGAGCTTTATCATCACA 2324
Db 2214 CTGCCAGGACGATCATATCCCCCAGAGGTGACAGGCTTCAAGGCGGGGTTCTCTCACA 2273
QY 2325 ATCCATCATTTCTGGCAGATACATTTAGGAGTCAAGTGTGAAACCAACATTTGAAATAT 2384
Db 2274 GCTGTGACGCTTGTGATCGACTTATTTGGGTGACAAACAGATATGACTCTGGATGGAAT 2333
QY 2385 TAATATAAAGAGATATTTCCAGTTTCCAAACAGGCAAAATACCAGATGTGCATTTCTT 2444
Db 2334 CACCTCCACGCTGAATTTTCCACCTGGATTCCTTGGGAATACCGGAGCTGTCTTCT 2393
QY 2445 TTATAAGTCTTCTACAGCAACAACATCTTGTATTAATGGCCGATCAACTGCTGTGAAAT 2504
Db 2394 TTATAGTCCCAATGATGTGACCCAGTCTCGAGTCTCGGAGATCAACACCACTCCCGT 2453
QY 2505 GAGGTGTAATCCTACTAAATCTGGAGCAGGAGTGAATTTAGTGTCCCCAGCAAGTGCACG 2564
Db 2454 CAGGTGAGTCCACAGAAAAGTGTCCCTGGGAAGTTGTGCTGCTGCCAGGAAGTGTCTCAGA 2513
QY 2565 AGGTACCTGTGATGGTGTACGTTCTATTTCTGTGGGAGAGTGTGAAGCTTGTGCCCTCT 2624

Qy	1248	TT	CAGAT	TG	GA	CC	AA	GA	AT	GT	AG	AC	CA	TG	TC	CA	CG	GA	AG	CA	GG	AG	CG	CT	GC	AC	TT	TG	GC	TT	1301				
Db	1250	CT	CCA	TGG	---	CT	CA	GA	CT	T	AC	CG	CT	TG	CC	CT	TG	CC	CT	GA	GG	AG	CT	GA	AC	CT	GC	CT	TG	TG	GG	GA	TT	1306	
Qy	1308	TG	AA	T	A	A	A	T	T	G	T	G	G	A	A	T	G	G	A	A	A	A	T	G	G	A	A	T	G	G	A	A	T	1367	
Db	1307	TG	AA	T	A	A	T	T	G	T	G	G	A	A	A	T	G	G	A	A	A	A	T	G	G	A	A	T	G	G	A	A	T	1366	
Qy	1368	GA	AT	T	CA	AA	T	G	CG	AT	GA	AT	GA	AT	T	G	G	AG	T	G	CG	AT	G	G	CT	GG	AG	AT	T	CC	AG	AT	1427		
Db	1367	CA	AC	T	CG	AG	T	CA	AA	GG	CA	T	CA	AG	CT	T	G	G	AG	T	G	CG	AT	G	G	CT	GG	AG	AT	T	CA	CA	CT	1426	
Qy	1428	GG	CT	GG	AG	T	CT	GA	CA	AT	G	AT	T	AA	CT	T	AA	CT	T	GA	CA	T	AT	T	GC	AT	T	CC	AG	AT	T	AA	CC	1487	
Db	1427	TG	CT	GG	AG	CT	CA	GA	CA	T	GA	CT	T	AA	CT	T	AA	CT	T	GA	CA	T	AT	T	GC	AT	T	CC	AG	AT	T	AA	CC	1486	
Qy	1488	AC	CA	AC	AT	CT	AT	GA	CT	GG	AG	CC	AC	---	---	GG	GT	CT	GA	AC	T	AG	GA	GA	AT	T	AA	CA	AT	T	T	GT	CT	1544	
Db	1487	TC	CG	CA	GT	CG	CA	T	GC	AG	AC	AG	CA	GA	GA	AT	AA	AG	GT	GC	CA	GA	AT	T	CA	CA	AT	T	CA	CA	AT	T	GT	CT	1546
Qy	1545	TG	AG	AC	CC	T	CT	GT	T	CA	CT	T	GA	CT	T	GT	T	GA	CT	T	CA	CT	T	GA	CT	T	GT	T	CA	CT	T	GA	CT	1604	
Db	1547	TG	AG	AC	CC	T	CT	GT	T	CA	CT	T	GA	CT	T	GT	T	GA	CT	T	CA	CT	T	GA	CT	T	GT	T	CA	CT	T	GA	CT	1606	
Qy	1605	T	A	CA	AA	T	G	T	G	G	T	G	G	T	G	G	A	CC	AA	GA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	1664	
Db	1607	CA	AC	CA	CT	CT	GT	G	AG	AC	GT	G	GA	AG	TT	T	CC	AA	AG	CA	AA	CA	AG	CT	T	AT	AC	CT	CA	CA	CT	CA	CT	1666	
Qy	1665	CT	T	CA	GA	AT	T	CA	CT	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	1724		
Db	1667	TG	AG	GA	GA	CA	CT	T	AC	CA	AG	CT	T	T	CA	CT	T	T	CA	CT	T	T	CA	CT	T	T	CA	CT	T	T	CA	CT	1726		
Qy	1725	T	A	A	T	A	G	A	G	CG	T	T	CA	CA	T	G	A	T	G	T	G	A	A	G	A	T	T	T	T	T	T	T	T	1784	
Db	1727	A	G	CA	G	GA	AG	T	AC	CA	CA	N	T	G	CG	AT	T	CT	CC	CA	AT	T	CT	CC	CA	AT	T	CT	CC	CA	AT	T	CT	1786	
Qy	1785	T	G	A	T	G	G	G	T	G	G	CT	CA	T	GC	CG	T	G	CT	GC	CG	T	G	CT	GC	CG	T	T	T	T	T	T	T	1844	
Db	1787	GA	AT	T	G	G	CT	G	G	CT	T	CA	CT	T	GC	CG	T	G	CT	GC	CG	T	T	CA	CT	T	GC	CG	T	G	CT				

[illegible]

Db 267 GCTTCATGCTCCAAAGAGCTGAGTACCACTATGAGTACACGGCGTGTGACAGCACGGG 326
Qy 231 CTCAGGTGGAGAGTGGCCATTCACAAATCTGCACTGAGCTGCTCGCCCTGCCTGACCC 290
Db 327 TTCAGGTGGAGGTGCGCGTCCGCATACCCCGGCGCTGTGCACACAGCGCTGTGACCC 386
Qy 291 AGTGAGAGCAAGAATGCATCTTCTCCTGTGCTTCTTGAGAGATCTAGAAATGAAGA 350
Db 387 CGTCAAGGCGCACCGAGTCTCTCTCTGCAACCCCGGGAGTTTCTGGATATGAAGA 446
Qy 351 CCAGGTATGCAGTGAAGTGGTGAAGGCACCTATTCCTTGGCCAGTGGCATCAAAATTTGA 410
Db 447 CGAGTCTGTGAAGCCATCGGTGAGGCGGTACTCCCTCGGCACAGGCATTCGGTTTTGA 506
Qy 411 TGAATGGGATGAATGCGCGGAGGATTTCTTAACATCCAACTTCATCGACACTGTGGT 470
Db 507 TGAGTGGATGAGTGGCCCATGGCTTTGCCAGCGCTTCAGCCCAACATGGAGCTGGATGA 566
Qy 471 GGGCCCTTCTGACAGCAGGCGAGCGGTGTAAACACTTCTTCTTGGAATCCCTCGTGGAAA 530
Db 567 CAGTCTGCTGAGTCCA---CGGGAAGTGTACTTCGTCCAAGTGGTTCCCGGGCGCA 623
Qy 531 CTACATAGAATCTAATCGTGAAGTGCAGCGTGTCTTTGATCTATGCTGTGCACCTTAA 590
Db 624 CTACATCGCTCCCAACAGGAGGAATGACAGCCACACTGATGTACGCGCTCAACCTGAA 683
Qy 591 GAAGTGAGGCTATGCTTCTTTGATGACAGTATGTGCAGCAACACATCTTCTTTGAGTT 650
Db 684 GCAATCTGCGCCGTTAACTTCGAATACTACTATCCAGACTCCAGCATCATCTTTGAGTT 743
Qy 651 CTTTATTCAAATGATGATGCGCAGGAGATGGACACCACTGACAAAGTGGGTAAACT 710
Db 744 TTTCTGTCAGAAATGACCACTGCGAGCCCAATGCAGATGACTC---CAGGTGGATGAAGAC 800
Qy 711 TACAGACAATGAGAAATGGGCTCTCATCTCTGATGCTGAATCAGGACGACAAACATACT 770
Db 801 CACAGAAA---AGGATGGAAATCCACAGTGTGGAGTAAATCGAGCAATAATGTGCT 857
Qy 771 CTACTGGAGAACTACAGCATCTCTTATGGTGTCTAAGCGGTCAAGCGCTGTCTGTTAAA 830
Db 858 CTATTGGAGAACCAAGCTTCTCAGTATGGACCAAGTACCCAGCGCTGTGCTGGTGA 917
Qy 831 AAATATCAATTTGAAGGGTGGGTACACATCAGAAATGTTTTCTTGCAGCCAGGCAC 890
Db 918 AAACATTTGCCATTAACAGGGTGGGTACACTTCAAGATGCTTCCCTGCAAACTGGCAC 977
Qy 891 ATTCAGCAACAAACAGCTTCATTCACACTGCCAGGTGTCCAGAAACACTTATTCGA 950
Db 978 GTATGCAGACAGCAGGCTCCTCTTTCTGCAAACTTTGCCAGCCAACTCTTATTCAAA 1037
Qy 951 GAAAGGACCAAGAAATGATAAGGTGTAAA---GACGACTCTCAATTTTCAGGATCCAG 1007
Db 1038 TAAAGGAGAACTTCTGCCACCACTGTGACCTTGACAAATATCAGAGAAAGGATCTTC 1097
Qy 1008 TGAGTGTACAGAGCCCTCCTGTACACAAAGACTATTTCCAGATCCATATCTCCATG 1067
Db 1098 TTTCTGTAACTGCGCCAGCTGTGCACAGAAAGATATTTCTACACACACAGCGCCTG 1157
Qy 1068 TGATGAAGAAGAAACACAGATTAATCTACAAGTGGATAGAGCCCAAAATCTGCCGGA 1127
Db 1158 CGATGCCAACGGAGACACAACTATGATAAATGGGCGAAGCGGAAATCTGTAGGGA 1217
Qy 1128 GGATCTACAGATGCTATTAGATTTGCCCTTCTGGAGAGAAAGAGTGTGCCCTTG 1187
Db 1218 GGACCTTGAGGGGCGAGTGAAGCTGCCTGCTGCTGGTGTGAAGACCACTGCCACCTG 1277
Qy 1188 CAACCTCGATTTTATACAAATGGATCATCTTCTTGCCATCCCTGTCTCTCGTGAACAT 1247
Db 1278 CAACCCAGGCTTCTTCAAAAACCAACAGCAGCCTGCCAGCCCTGCCCATATGGTTCTTA 1337
Qy 1248 TTCAGATGGAACCAAGAAATGTAGACCATGTCCAGCAGGACGAGGCTGACACTTGGCTT 1307
Db 1338 CTCCAATGG---CTCAGAGTGTACCCGCTGCTGACAGGACTGAACCTGCTGTGGGAT 1394

Qy 1308 TGAATATAAATGGTGGAAATGCTCTTCTCGGCAACATGAAAACTTCTGCTTCAATTTGG 1367
Db 1395 TGAATACAAATGGTGGAAACACGCTGCCCAACAAATGGAACACGCGTTCTCAGTGGGAT 1454
Qy 1368 GAATTCAAAGTTCGATGGAATGAATGGTGGAGGTGGCTGGAGATCATATCCAGAGTGG 1427
Db 1455 CAACTTCGAGTACAAAGGCATGACAGGCTGGGAGTGGCTGGTGTACATATACACAGC 1514
Qy 1428 GGCTGGAGTTCGACATGATGATCTGATCTTAACCTTGAATATCCAGAGATTTAAACC 1487
Db 1515 TGCTGGAGCCTCAGACAAATGACTTTCATGATCTCTGCTGTGGTGGCCAGGATTTAGACC 1574
Qy 1488 ACCAACATCTATGACTGGAGCCAC---GGGTTCTGAACCTAGGAAGAATAACATTTGCTTT 1544
Db 1575 TCCGAGCTCGGTGATGGGACACAGAGAAATGAAGGTGGCCAGAAATCACAATTTGTCTTT 1634
Qy 1545 TGAGACCTCTGTTCCAGCTGACTGTGTTTGTACTTTCATGTTGGATATTAATAGAAAAAG 1604
Db 1635 TGAGACCTCTGTTCTGTAACCTGTGAGCTTACTTTCATGGTGGTGTGAATTTCTAGGAC 1694
Qy 1605 TACAATGTGTAGAAATCGTGGGTGGAACCAACAAAGAAACAAAGCTTACACCATATCAT 1664
Db 1695 CAACACTCCTGTGGAGAGCTGGAAAGGTTCCAAAGGCAAAACAGTCTATATACCTATCAT 1754
Qy 1665 CTTCAAGAATGCAACTTTTACATTTACATGGGCATTTCACAGAACTTAATCAGGGTCAAGA 1724
Db 1755 TGAGGAGAACACTACACGAGCTTCACCTGGGCTTCACAGAGCACACTTTTCATGAGGC 1814
Qy 1725 TAATAGACGGTTCATCAATGACATGGTGAAGATTTATCTATCACAGCCACTAATGCACT 1784
Db 1815 AAGCAGGAATACACCAATGAGCTTGCCAAAGATCTACTCCATCAATGTCAACCAATGTTAT 1874
Qy 1785 TGATGGGTGGCGTCTCATGCCGTGCCCTGTCGCCCTGTGAAAGCTCTGATGTGGGCTCTC 1844
Db 1875 GAATGGCTGGCTCTCTACTGCGCTGCCCTGTGAAAGCTCTGATGTGGGCTCTCTC 1934
Qy 1845 GTGTGTCCTCCCTCCAGGCCACTTACATTCAGAAAGAAACCAACAGCTGCAAGGAATG 1904
Db 1935 CTGCACCTCTGCTGCTGCTGTTACTATATTCACGAGATTCAGGAACCTGCCACTCTG 1994
Qy 1905 TCCACCTGACACTACCTGTCCATACATCAGTCTATGSCAAAGAGGCTTGTATTCACATG 1964
Db 1995 CCCCCTTAACAACTTCTGAAAGCCCAACAGCTTATGGTGTCCAGCGCTGTGTGCCCTG 2054
Qy 1965 CGGCTGGGAGTAAACAAATCAGACCAATTCGGTTTGTCTATAGTACTGCTTTTCTA 2024
Db 2055 TGGTCCAGGGACCAAGAACACAAAGATCCACTCTCTGCTACAATGATTCACCTCTC 2114
Qy 2025 CCATGAAAAAGAAATCAGATTTTGCACCTATGACTTTAGCAACCTCAGCAGTGTGGGCTC 2084
Db 2115 AGGCAACACTCCAACAGGACTTTCACATCAACTTCTCCGCTTTGGCAACACCGTCA 2174
Qy 2085 ATTAATGAATGCCCGCAGCTTCACCTCCAAAGAAACAAATACTTCCATTTCTCAATAT 2144
Db 2175 TCTTCTGGAGGGCCAAAGCTTCACTTCCAAAGGGTTGAAATACTTCCATCACTTTACCT 2234
Qy 2145 CAGTTTATCTGGCATGAGGGAAGAAGTGGCTCTCTGTACCAACAATATACAGACTT 2204
Db 2235 CAGTCTCTGTGAAACACAGGAGTGAAGAAATGCTGTGTGACCCGCAATGTCTACTGACT 2294
Qy 2205 TACAGTAAAGAAATAGTGGCAGGCTCAGATGATTACACAAATTTGGTAGGGGCAATTTGT 2264
Db 2295 CCGGATTCCTGAGGGTGAGTCAGGG-----TTCCTCAATCTATCACAGCTACGT 2345
Qy 2265 ATGCCAGTCAACAATATTTCTTCTGAAAGTAAAGGGTTTCCGAGCAGGCTTATCATCACA 2324
Db 2346 CTGCCAGCAGTCACTATCCCGCCAGAGGTGACAGGCTACAGGCGCGGGTTTCTCTACA 2405
Qy 2325 ATCCATCATCTTGGCAGATACATTTATAGGAGTTCAGAGTTGAACCAATTTGAAATAT 2384
Db 2406 GCCTGTGAGCTTGTGTCGACTTATTGGGTGACACAGATATGACTCTCTGGATGGAAT 2465

Qy	2385	TAATATAAAAGAGATATGTTCCAGTCTCCACAAGCCAAATACCAGATGTGCATTTCTT	2444
Db	2466	CACCTCCCCAGCTGAACCTTTTCCACCTGGAGTCTCTGGGAATACCGGACGTGATCTCTT	2525
Qy	2445	TTATAAGTCTTACACACAACATCTTGTATTAAATGGCCGATCACTGCTGTGAAAT	2504
Db	2526	TTATAGTCCAAATGATGTGACCCAGTCTCGAGTCTGGGAGATCAACCACTCCCGCT	2585
Qy	2505	GAGGTGTAACTCTACTAAATCTGGAGCAGGATGATTTCAGTCCCAAGCAAGTGCCGACG	2564
Db	2586	CAGGTGAGTCCACAGAAAATGTCCCTGGAATTTGCTGCTGCCAGAGAGTGTCTCAGA	2645
Qy	2565	AGGTACCTGTGATGGGTGTACGTTCTATTTCTGTGGGAGAGTCTGAAGCTTGTGCCCTCT	2624
Db	2646	TGGGACCTGTGATGGGTGCAACTTCCACTTCTGTGGGAGAGCGCGCTGCTGCTGCCGCT	2705
Qy	2625	GTGTACGGAGCATGACTTCCATGAGATTTAGGGAGCGCTGCAGAGAGAGATTTCAAGAAAC	2684
Db	2706	CTGCTCAGTGGCTGACTACCATGCTATTCGTGACGAGCTGTGTGGCTGGGATCCAGANGAC	2765
Qy	2685	CTTGTATGCTGGAAATGAACCTTAATGGTGCATTAAGGAATTTCTTTGCTGCTGAGAAAAA	2744
Db	2766	TACTTACGTGNGCAGAACCAAGCTATGCTCTGGTGGCATTTCTCTGCTGTGAGCAGAG	2825
Qy	2745	GTTGGCAACCTGTGAACGGTTGACTTTTGGCTGAAGGTGGGAGCGGTGTGGGAGCTTT	2804
Db	2826	AGTCACCATCTGCAAAACCATAGATTTCTGGCTGAAAGTGGGCATCTCTGCAGGCACCTG	2885
Qy	2805	TACTGCCGTTTGTGGTGGCTCTGACCTGCTACTTCTGGAAGAAAGATCAAGAGAAAAA	2864
Db	2886	TACTGCCATCTGCTACCGCTTCTGAGCTGCTACTTTTGGAAAAAGAAATCAAAACCTAGA	2945
Qy	2865	GAAGACCATTTTGAATCTGTCTCA	2887
Db	2946	GTACAAGTACTCCAGCTGGTGA	2968
RESULT 7			
AX127742			
LOCUS	AX127742	2610 bp	DNA linear PAT 15-MAY-2001
DEFINITION	Sequence 17 from Patent WO0131003.		
ACCESSION	AX127742		
VERSION	AX127742.1	GI:14134389	
KEYWORDS	human.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 2610)		
TITLE	Delneste,Y., Magistrelli,G., Jeannin,P. and Bonnefoy,J.Y.		
JOURNAL	Cloning, expression and characterisation of a gene expressed in tumour cells and involved in the regulation of the immune response		
FEATURES	Patent: WO 0131003-A 17 03-MAY-2001; PIERRE FABRE MEDICAMENT (FR)		
SOURCE	Location/Qualifiers		
	1..2610		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
CDS	<1..>2610		
	/codon_start=1		
	/note="unnamed protein product"		
	/protein_id="CAC38913.1"		
	/db_xref="GI:14134390"		
	/translation="TGPDLHAKSEYHYEYTAACDSTGSRWRVAVPHPTGCTSLPDP		
	VKGTECSFSCNAGEFLDMKDOCKPCAEGRYSLSGTGRFDEWDELPHGFASLSANMEL		
	DDSAESTGNTSSKWPGRDYLASNTDECTATLMYAVNLKQGTVNFYYYPDSSII		
	FEFVNDQCPQNAEDSRMKMTTEKGEFHSVELNRGNVLYWRMTAFSVMTKVPKYV		
	LVNRNLTJGVAYTSECCFPCKGTADKQSSFKLCPANSYSNKGTFSCHQCPDKYS		
	EKGSSSCNVPACTDKDYFYTHTACDANGETOLMYKWKPKICSEDLGAKLPASV		
	KTTCPCNQPFRTKMTGTCQCPPIGYSVNGSDCTRCRCPAGTEAVGFYKKNWTLPTNN		
	ETTVLSINFEYKGNMGEVAGDHIYTAAGASDNFMILLTVVPFPPQPSVMTDTN		
	KEVARITFFETLCSVNCLEYFMVGNSTRNTPETWKGSKGOSYTYIIIEENTTYS		
	TWAFORTTFHEASRKYTNDAKIYSINVTWMNGVASYRCPCALEADSVGSSCTSCPA		


```

Db 1983 ACGCAACACTCCAAACAGGACTTTCAACTACAACTTCTCGCTTTGGCAAAACCGTCAC 2042
Qy 2085 ATTAAATGATGCCCCAGCTTCACCTCCAAAGGAAACAAATACCTTCTTCATAT 2144
Db 2043 TCTTCTGAGGGGCAAGCTTCACCTCCAAAGGGTTGAAATACCTTACCTTACCCCT 2102
Qy 2145 CAGTTATGTGGGCATGAGGGAAGAAGTGGCTCTCTGTACCAACAATATAACAGACTT 2204
Db 2103 CAGTCTGTGGAACACAGGAGTGAAGAAATGCTGTGACCCAGCAATGTCACTGACCT 2162
Qy 2205 TACAGTAAAGAAATAGTGGCAGGGTCAGATGATACACAATTTGGTAGGGCAATTTGT 2264
Db 2163 CCGGATTCCTGAGGGTGAAGTGGT-----CTCCAAATCTATCACAGCCTACGT 2213
Qy 2265 ATGCCAGTCAACAATATTCTCTCTGAAAGTAAGGGTTTCGAGCAGCCTTATCATCACA 2324
Db 2214 CTGCCAGGCACTATCATCCCCCAGAGGTGACAGGCTACAGGCGGGGTTTCTCTCACA 2273
Qy 2325 ATCCATCATTTCTGCGAGATACATTCATAGGAGTCAAGTGTGAACACCACTTGAATAAT 2384
Db 2274 CCTGTGACGCTTCTGATGACCTATTGTTGGGTGACACAGATGATGACTCTGGATGGAAT 2333
Qy 2385 TAATATAAAGAAATAGTTCACAGTTTCCCAAGTCCAAAGCAAGCAATACAGATGTCATTTCT 2444
Db 2334 CACCTCCCAAGCTGAACCTTTTCCACCTGGAGTCCCTTGGGAATACCGGACGTGATCTTCT 2393
Qy 2445 TTATAGTCTTCTACAGCAACAACATCTTGTATTAATGGCCGATCAACTGCTGTGMAAT 2504
Db 2394 TTATAGTGTCAATGATGTGACCCAGTCTGCGAGTCTGGGAGATCAACCCATCCGCGT 2453
Qy 2505 GAGGTGTAATCTACTAATCTGGAGCAGGAGTATTCAGTCCCAAGCAAGTGCAGC 2564
Db 2454 CAGGTGCAGTCCACAGAAACTGCTCCGGAAGTTGCTGCTGCCAGCAAGTCTCAGA 2513
Qy 2565 AGTACCTGTGATGGGTGATGCTTCTATTTCTGTGGGAGAGTGTGAAGCTTGGCCCTCT 2624
Db 2514 TGGGACCTGTGATGCTGCAACTTCCACTCTGTGGGAGAGCGGCTGTGTCGCCCT 2573
Qy 2625 GTGTACGAGGATGACTTCCATGAGATTGAGGAGCCTTCAAGAGAGGATTTCAAGAAAC 2684
Db 2574 CTGCTCAGTGGCTGACTACCGTGTATCGTCAGCAGCTGTGTGGCTGGATCCAGAAC 2633
Qy 2685 CTGTATGTGGGAATGAACCTAAATGTGTCATTAAGGAATTTCTTTCCTCGAGAAAA 2744
Db 2634 TACTTACGTGTGGGAGAACCAAGCTATGCTGCTGGTGGCATTTCTCTGCTGACGACAG 2693
Qy 2745 GTTGGCAACCTGTGAACGGTTGACTTTTGGCTGAAGTG 2784
Db 2694 AGTCACCATCTGCAAAACCATAGATTCTTGGCTGAAGTG 2733

RESULT 9
AK057647
LOCUS AK057647 2895 bp mRNA linear PRI 01-AUG-2002
DEFINITION Homo sapiens CDNA FLJ33085 fis, clone TRACH2000420, weakly similar
to MAJOR SURFACE-LABELED TROPHOZOITE ANTIGEN PRECURSOR.
ACCESSION AK057647
VERSION AK057647.1 GI:16553406
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens trachea cDNA to mRNA, clone_lib:TRACH2
clone:TRACH2000420.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS Suzuki,O., Sasaki,N., Aotsuka,S., Shoji,T., Ichihara,T.,
Shohata,N., Matsumoto,K., Hirano,M., Sano,S., Nomura,R.,
Yoshikawa,Y., Matsumura,Y., Moriya,S., Chiba,E., Momiyama,H.,
Onogawa,S., Kaeriyama,S., Satoh,N., Matsunawa,H., Takahashi,E.,
Kataoka,R., Kuga,N., Kuroda,A., Satoh,I., Kamata,K., Takami,S.,
Terashima,Y., Watanabe,M., Sugiyama,T., Irie,R., Otsuki,T.,
Sato,H., Ota,T., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,

```

```

Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
Wagatsuna,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,X., Sugano,S.,
Nagahara,I.K., Masuno,Y., Nagai,K. and Isogai,T.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2895)
Isogai,T., Otsuki,T. and Sugiyama,T.
Direct Submission
Submitted (24-Oct-2001) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 252-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing;
Research Association for Biotechnology (RAB); cDNA library
construction; Helix Research Institute (HRI) (supported by Japan
key technology Center etc.); 5'- & 3'-end one pass sequencing; RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing; RAB and
HRI.
FEATURES
Source
1..2895
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="TRACH2000420"
/tissue_type="trachea"
/clone_lib="TRACH2"
/note="cloning vector: pME18SFL3"
BASE COUNT 692 a 825 c 714 g 664 t
ORIGIN
Query Match 20.9% Score 709.6; DB 9; Length 2895;
Best Local Similarity 57.0%; Pred. No. 2.7e-155;
Matches 1550; Conservative 0; Mismatches 999; Indels 171; Gaps 7;
Qy 171 GCTTCTCTCTTCCAGGAGAAAGATATCATCTTTGAATATACGGAATGTGTATAGCAGTGG 230
Db 214 GCTTCATGCCGCAAGAGTCTGATACCATGATGATGATGATGATGATGATGATGATGATG 273
Qy 231 CTCAGGTGGAGAGTGGCATTCCTCAAAATTCGTGAGTGAGTCTCTGCTGCTGCTGCTGCTG 290
Db 274 TTCAGGTGGAGAGTGGCGGCGCGCATACCCCGGCGCTGTGCACGAGCTGCTGCTGCTGCTG 333
Qy 291 AGTGAGAGGCAAGAATGCACCTTTCTCTGTGCTTCTGGAGAGTATCTAGAAATCAAGAA 350
Db 334 CGTCAAGGGCACCGAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 393
Qy 351 CCAGTATGCAAGTATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 410
Db 394 CCAGTATGTAAGCATGCTGCTGAGGCGCGCTACTCTCTCTCTCTCTCTCTCTCTCTCTCT 453
Qy 411 TGAATGGGATGAATTTGCGCGGAGGATTTTCTAACAATCGCAACATTCATGGACATCTGCT 470
Db 454 TGAGTGGGATGAGTGTGCGCCCATGCGCTTTGCCAGCGCTCTCAGCAACATGGAGTGGATGA 513
Qy 471 GGGCCCTTCTCAGACAGCGCCAGCGCTGTAAACAACATCTTCTTGGATCCCTCGCTGGAAA 530
Db 514 CAGTCTCTCTGAGTCCA---CGGGAACCTGTACTTCTGCTCAAGTGGTTCCTCCGCGGCGA 570
Qy 531 CTACATAGAATCTAATCTGTGATGCTGACCGGTGTCTTTGATCTATGCTGTGTCACCTTAA 590
Db 571 CTACATCGCTCCACACAGGAGGATGACAGCCACACTGATGTACGCGCTCAACCTGAA 630
Qy 591 GAAGTCAGGCTATGCTTCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 650
Db 631 GCAATCTGGCAGCGGTAACTTCCGAATCTACTATCTCCAGCATCTCCAGCATCTCTTTGAGTT 690
Qy 651 CTTTATTTCAAAATGATCAGTGTCCAGGAGATGCACACCACTGACAAAGTGGGTAAAACT 710
Db 691 TTTCTGTCAGATGACAGTGTCCAGCCCAATCCAGATGACTC---CAGGTGATGAAGAC 747
Qy 711 TACAGACAATGAGAAATGGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 770

```

Db 748 CACAGAGAA---AGGATGGGAATTCACAGTGTGGAGCTAAATCGAGGCAATAATGTCTC 804
Qy 771 CTAATGGAGAACTACAGGCATCTTATGGTTC TAAGCGGTCAAGCGTGTCTGTGTA 830
Db 805 CTAATGGAGAACACACGCTTCTCAGTATGGACCAAGTACCCAGCCCTGTCTGTGTG 864
Qy 831 AATATCAAAATTAAGGGGTGGGTACACATCAGAAATGTTTCTTGCAGCCAGGCAC 890
Db 865 AAACATTCGCATAACAGAG----- 883
Qy 891 ATTACAGCAACAAACCGAGTTTCAATCACTGCCAGGTGTCTCCAGAAACACCTATTCTGA 950
Db 884 ----- 883
Qy 951 GAAAGAGCCAAAGAAATGTATAAGGTGTAAAGACGACTCTCAATTTTCAGGATCCAGTGA 1010
Db 884 -----AAAGGATCTTCTTC 897
Qy 1011 GTGTACAGAGCGCTTCCCTGTACACAAAGACTATTTCCAGATCCATPCTCCATGTGA 1070
Db 898 CTGTAACTGTGCCCGCAGCTTGCACAGACAAAGATTATTTCTACACACACACCGCTGGCA 957
Qy 1071 TGAAGAAGGAAAGACACAGATAATGTACAAAGTGGATAGAGCCCAAAATCTGCCGGAGGA 1130
Db 958 TGCCACGGGAGAGACACAACACTCATGTACAAATGGGCCAAGCCGAAATCTGTAGCGAGGA 1017
Qy 1131 TCTCAGAGATGCTATTAGATTGCCCTTCTCGAGAGAGAAAGGATGTGCCCTTGC 1190
Db 1018 CCTTGAGGGGCGAGTGAAGCTGCCCTCTCTGTGTGAAGACCCACTGCCACCTGCCAA 1077
Qy 1191 CCCTGGATTTTATACAATGGATCATCTTCTTGGCAATCCCTGTCTCTGGAACATTTTC 1250
Db 1078 CCGAGGCTTCTTCAAAACCAACACAGCAGCACTTGCAGCCCTGCCCATATGTGTCTACTC 1137
Qy 1251 AGATGAACCAAGAAATGTAGACCATGTCCAGCAGGAAGAGGCTGCACATTTGGCTTGA 1310
Db 1138 CAATGG---CTCAGACTGTACCCGCTGCCCTGCAGGACTGACCTGTGTGGGATTTGA 1194
Qy 1311 ATATAATGGTGGATGTCTTCTTCCGCAACATGAAATCTTCTCTCTCAATGTTGGAA 1370
Db 1195 ATACAAATGGTGGAGACAGCTGCCACACAAACATGAAACGACCGTCTCAGTGGGATCAA 1254
Qy 1371 TTCAAAGTTCGATGAATGAATGGTGGAGGTGGCTGGAGATCATATCCAGATGGGGC 1430
Db 1255 CTTGAGTACAAGGCTATGACAGGCTGGAGGTGGCTGGTATCATATTAACAGCTGC 1314
Qy 1431 TGGAGGTTCTGACAAATGATTTACTCTTAACTTTGATATCCAGGATTTAAACACACC 1490
Db 1315 TGGAGCCTCAGACAAATGACTTCATGATCTCTACTCTGTTGTGCCAGATTTAGACCTCC 1374
Qy 1491 ACATCTATGACTGGAGCCAC---GGTTCGAACTAGAGAAATTAACATTTGTCTTTGA 1547
Db 1375 GGAGTCGGTGTGGGAGACACAGAAATTAAGAGGTGGCCAGAAATCACATTTGTCTTTGA 1434
Qy 1548 GACCCTCTGTTCAGCTGACTGTGTTTGTACTTTCATGTGGATATTAATAGAAAAGTAC 1607
Db 1435 GACCCTCTGTCTGTAAGTGTGAGCTCTACTTTCATGTGGGTGGTGAATCTTAGACCAA 1494
Qy 1608 AATGTGTAGAAATCTGGGTGGAAACCAAGAAAGAAATTAACACCATATCATCTT 1667
Db 1495 CACTCCTGTGGAGACGTGGAAAGTTCCAAAGGCAACAGCTTATACCTACATCATTTGA 1554
Qy 1668 CAAGAATGCAACTTTTACATTTACATGGGCATTCAGAGAACTAATCAGGGTCAAGATAA 1727
Db 1555 GGAGAACACTACACAGAGCTTCACTCTGGCCCTTCCAGAGGACCACTTTTCATGAGGCAAG 1614
Qy 1728 TAGACGGTTTCATCAATGACATGTGCAAGATTATTTCTATCACAGCCACTAATGACAGTTGA 1787
Db 1615 CAGGAAGTACACCAATGACGTGGCCAGATCTACTTCCATCAGTGTACCAATGTTATGAA 1674
Qy 1788 TGGGTGGCGCTCCTCATGCCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1847
Db 1675 TGGCGTGGCTCCTACTGCCGTGCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1734

Qy 1848 TGTCCCTCTCCCTCCAGGCCACTACATTTGAGAAAGAAACCAACCAAGTGCAGGAATGTCC 1907
Db 1735 CACCTCTTGTCTCTGTGTACTATATTGACCGAGATTCAGGAACCTGCCACTCTCTGCC 1794
Qy 1908 ACCTGACACCTACCTTGTCTCATACATCAGTGTCTATGGCAAGAGGCTTGTATTTCCATGGG 1967
Db 1795 CCCTTAACAAATTTCTGAAAGCCACACGCTTATGTGTCCAGGCTGTGTGCCCTGTGG 1854
Qy 1968 GCCTGGGAGTAAACAAATCAGGACCATTTGGTTCATAGTGACTGCTTTTCTATACA 2027
Db 1855 TCCAGGGACCAAGAAACAAAGATCCACTCTCTGTGTGTACAAATGATTCACCTTCTCAG 1914
Qy 2028 TGAAGAAAGAAATCAGATTTTGTCACTATTTAGCAACCTCAGCAGTGTGGGCTCAT 2087
Db 1915 CAACACTCCACACGAGACTTTCAACTACAACTTCTCCCTTTGGCAACACCGTCATCT 1974
Qy 2088 AATGAATGGCCCTCAGCTTCACTCCAAAGAAACAAATACTTCCATTTCTCAATATCAG 2147
Db 1975 TGTGGAGGGCCAGCTTCACTTCCAAAGGTGTGAATACCTTCCATCACTTTACCTCAG 2034
Qy 2148 TTTATGTGGCATGAGGGGAGAGAGATGGCTCTCTGTGTACCAACAAATATAACAGACTTTAC 2207
Db 2035 TCTCTGTGAAACACAGGCTAGGAAATGTCTGTGTGCAACCAATATGTCTGACCTCCG 2094
Qy 2208 AGTAAAGAAATAGTGGCGGCTCAGATGATTACAAATTTGGTGGGCAATTTGTATG 2267
Db 2095 GATTCCTCAGGCTGAGTCAGG-----TCTCCAAATCTATCACAGGCTACGCTG 2145
Qy 2268 CAGTCAACAAATATCTCTTCTGAAAGTAAAGGTTCGAGCAGCGCTTATCATCAACAATC 2327
Db 2146 CAGGCACTCATCATCCCGCCAGAGGTGACAGGCTACAAAGCGCGGGTTTCTCTCACAGCC 2205
Qy 2328 CATCAATCTGCGCAGATACATTCATAGGATGACAGTGTGAAACCAATTTGAAATATTA 2387
Db 2206 TGTACAGCTTCTGATCGACTTATTTGGGTGACAAACAGATATGACTCTGGATGGAATCAC 2265
Qy 2388 TATAAAGAAATATGTTCCAGCTTCCAAACAGCAATATACAGATGTGCATTTCTTTTA 2447
Db 2266 CTCCCACTGAACTTTTCCACCTGGAGTCTTGGGAATACCGGACGATCTCTTTTA 2325
Qy 2448 TAGTCTCTACAGCAACAACTCTTGTATTAATGGCGGATCAACTGCTGTGAAATGAG 2507
Db 2326 TAGTCCAATGATGTGACCCAGTCTGCACTTCTGGAGATCAACCACTCCGCTCAG 2385
Qy 2508 GTGTAATCTACTAAATCTGGAGCAGGATGATTTTCACTCCCAAGTGCAGCAGG 2567
Db 2386 GTGAGTCCACAGAAACTGTCCCTGGAGATTTGCTGCTGCCAGGACGCTCTCAGATGG 2445
Qy 2568 TACCTGTGATGGGTGTAGTCTATTTCTGTGGGAGTGTCTGAAAGCTTCCCTCTCTG 2627
Db 2446 GACCTGTGATGGTGTCAACTTTCCACTTCTGTGGAGAGCGGCTGCTTCCCGCTCTG 2505
Qy 2628 TACGAGCATGACITTCATGAGATTTGAGGAGCCTGCAAGAGAGATTTTCAGGAACCTT 2687
Db 2506 CTCAGTGGCTGAATACCATTCATCTGTCAGCAGCTGTGTGGCTGGATCCAGAGACTAC 2565
Qy 2688 GTATGTGTGGAAATGAACCTAAATGCTGATTTAAAGGAATTTCTTGTGCTGAGAAAGT 2747
Db 2566 TTACGTGTGGGAGAACCCAGCTATGCTCTGTGGCATTTCTGCTGCTGAGCAGAGAT 2625
Qy 2748 GGCACCTGTGAAACGGTGTGACTTTTGGCTGAAGTGGGAGCGGCTGTGGAGCTTTTAC 2807
Db 2626 CACCATCTGCAAAACCAATAGATTTCTGGCTCAAGTGGGCTCTCTGCAAGCAGCTGTAC 2685
Qy 2808 TGGCTTTTGTGCTGGCTCTGACCTCTGACCTCTCTGCAAGAAAGATCAAAAGAAAGAA 2867
Db 2686 TGGCTCTCTCAGCTGCTTGGCTGCTTGTGACCTGCTACTTTTGGAAAAAGAACTAGAGTA 2745
Qy 2868 CACCAATTTTGAATCTGTTC 2887
Db 2746 CAAGTACTCCAAAGCTGGTGA 2765

RESULT 10

BC031648

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

BC031648 2449 bp mRNA linear PRI 26-JUN-2002
Homo sapiens, KIAA1324 protein, clone MGC:35166 IMAGE:5169952,
mRNA, complete cds.

BC031648
MGC.

BC031648.1 GI:21594947

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 2449)

Direct Submission

Submitted (06-JUN-2002) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgapbs-re@mail.nih.gov

Tissue Procurement: Life Technologies, Inc.

cDNA Library Preparation: Life Technologies, Inc.

DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: amg@bcm.tmc.edu

Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,

Yoon, V.S., Kowls, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,

Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAK Plate: 51 Row: d Column: 23

This clone was selected for full length sequencing because it

passed the following selection criteria: Hexamer frequency ORF

analysis.

FEATURES

Source

Location/Qualifiers

1..2449

/organism="Homo sapiens"

/db_xref="LocusID:57535"

/db_xref="taxon:9606"

/clone="MGC:35166 IMAGE:5169952"

/tissue_type="Brain, adult medulla"

/clone_lib="NIH_MGC_119"

/lab_host="DH10B"

/note="vector: pCMV-SPORT6"

208..2199

/codon_start=1

/product="KIAA1324 protein"

/protein_id="AAH31648.1"

/db_xref="GI:21594948"

/translation="MYKAPKICSEDELEGAKLPASGVKTHCPNPGPFKTNSTC

OPCYGYSNGSDCTCPAGPEAFYKWNLTPTNMTTVLSGINFYKGMTGW

VAGDHVTAAGSANDFMILTLLVPPRPPOSVMADTENKEVAFITVFELCSVNC

LYFMVGNSTRNTVETWKSQKQSYIIIEENTTSFTWAFORTTFHEASRYTND

VAKIYSNVTNMVNGVASYCRPCALEASDVGSCTSPAGYIDRDSGTCHSCPVT

LKAHPYGVQACVPGPGTKNNKTHSLCYNDCTFSRNPTRTFNFSALANTVLAG

GPSTSLGKLVFHHFTLSLGNQGRKMSVCTDNVTLRIPEGESGSKSITAYVCOAV

IPPEYGVYKAGVSQPVSLADRLIGVTDMLGITSPALFHLSESLIPDVFYFR

SNVDYSCSSGRSTIVRCSPOKTPGSLLLPGTCSGDGCDGNGFHLFESAAACPL

CSVADYHAIYSSCVAGKQTYVMREPKLCSGGISLPEQRVITCKTIDFLWLVGISAG

FTCKLLTLVTCYFWKKNQKLEYKSLVMNATLKDCDLPADSCAIMEGEDVEDDLI

TCKSKSLFTSKRTKPTDGFDSVPLKTSSTSGGPDMDL"

618 a 690 c 563 g 578 t

BASE COUNT

Query Match

Best Local Similarity

17.7%; Score 600.2; DB 9; Length 2449;

59.5%; Pred. No. 1.2e-129;

Matches 1076; Conservative 0; Mismatches 718; Indels 15; Gaps 3;

QY 1082 AGACACAGATAATGATCAAGTGSATAGAGCCCAAAATCTGCCGGGAGGATCTCACAGATG 1141
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 197 AGACACAACTCATGTACAAATGGCCAAAGCCGAAATCTGTAGCGAGGACCTTTGAGGGG 256
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

QY 1142 CTATTAGATTGCCCTCTCTGGAGAGAAAGAGATTGTCCGGCTTTCACACCTCGGATTTT 1201
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 257 CAGTGAAGCTGCCTCTCTGTGTGAAGACCCACTGCCACCCCTGCAACCCAGGCTTCT 316
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

QY 1202 ATAAACATGGATCATCTCTTCCATCCCTCTCTCTGATGATTTTTCAGATGGAACCA 1261
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 317 TCAAAACCAACAAACAGCACTGCCACCTGCCCATATGTTCTTCTACTCCAATGG---CT 373
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

QY 1262 AAGATGTAGACCATGTCCAGCAGGAAGAGCGCTGCACCTTGGCTTTGATATAAATGCT 1321
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 374 CAGACTGTACCGCTGCCCTGCAGGACTGAACCTGCTGTGGGATTTGAATACAAATGTT 433
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

QY 1322 GGAATGTCTTCTTGGCAACATCAAACTTCTCTTCAATTTGGGAATTCAAAGTCGG 1381
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 434 GGAACAGCTGCCACACACATGGAACACCGTCTCTAGTGGATCACTTCGAGTACA 493
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

QY 1382 ATGGAATGAATGGTGGAGGTGGGTGGAGATCATATCCAGAGTGGGTGGAGTTCGTG 1441
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 494 AGGGATGACAGGCTGGGAGTGGGTGGTGTATCATATTTACACAGCTGCTGGAGCCTCAG 553
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

QY 1442 ACAATGATTACCTGATCTTAACTTCCATATCCAGGATTTAAACCAACCAACATCTATGA 1501
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 554 ACAATGACTTCATGATCTCTACTCTGGTTGTGCCAGGATTTAGACCTCGCGATCGGTGA 613
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

QY 1502 CTGGAGCCAC---GGGTCTCTGAACAGGAAGATTAACATTTGTTTGGAGACCTCTGTT 1558
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 614 TGGCAGACACAGAGAATAAGAGGTGGCCAGAAATCACATTTGCTTTGAGACCTCTGTT 673
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

QY 1559 CAGCTGACTGTGTTTGTACTTCTAGTGGATTAATAAGAAAAGTACAAATGTGGTAG 1618
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 674 CTGTGAACCTGTAGCTTCTACTTCTGTTGGTGTGTAATTTCTAGGACCAACACTCTCTGG 733
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

QY 1619 AATCTGGGGTGGAAACCAAGAAAGAAAGCTTACACCCATATCATCTTTCAAGATGCAA 1678
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 734 AGACGTGGAAAGTTCCTCAAGGCAACAGCTCTATACCTACATCATTTGAGGAGAACACTA 793
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

QY 1679 CTTTACATTTTACATGGGCATTCAGAGAACTAATCAGGGTCAAGATAATAGACGGTTCA 1738
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 794 CCACGAGCTTCACTTGGGCTTCCAGAGGACCACTTTTTCATGAGGCAAGCAAGTACA 853
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

QY 1739 TCAATGACATGTGAAGATTATTTATATACAGGCACTAATGAGTTGATGGGTGGCCT 1798
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 854 CCAATGACGTTGCCAAGATCTACTCCATCAATGTCCAAATGTTATGAATGGGTGGCCT 913
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

QY 1799 CTTATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1858
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 914 CTTATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 973
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

QY 1859 CTTCCAGGCACTACATTTGAGAAAGAAACCAACAGTGCAGGAATGTCACCTGACACCT 1918
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 974 CTGCTGGTTACTATATGACCGAGATTCAGAACTGCACTCTGCTGCTGCTGCTGCTGCT 1033
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

QY 1919 ACCTGTCCATACATCAGGTCTATGCAAGAGAGGCTTGTATTCATCGCGGCGCTGGAGTA 1978
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 1034 TTCTGAAAGCCCAACAGGCTTATGTTGTCCAGGCTGTGTGCTGCTGCTGCTGCTGCTGCT 1093
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

QY 1979 AAAACAATCAGGACCATTCGGTTTCTATAGTACTGCTTTTCTACCATGAAAGAA 2038
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 1094 AGAACACAAGATCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1153
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

QY 2039 ATCAGATTTTGCATATGACTTTAGCAACCTCAGCAGTGTGGGCTCATTAATGAATGGCC 2098
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 1154 CCAGGACTTCACTACACTTCTCGGCTTTGGCAACACCCGCTCACTCTTGTGTTGAGGCG 1213
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

QY 2099 CCAGCTTCACTCCAAAGAACAAAATCTTCCATTTCTTCAATATCAGTTTATGTGGCG 2158
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 1214 CAAGCTTCACTTCCAAAGGTTGAAATCTTCCATCACTTTTACCCTCACTCTCTGTGGAA 1273
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 630 CCTAAGACCTCTGATGTGGGCTCCTCTGACACCTTTGCTCTGCTGTTACTATATTG 689
QY 1877 AGAAGAAACCAACACAGTGCAGGAATGTCACCTCGACACCTACCTGTCCATACATCAGG 1936
Db 690 ACCGAGATTCAGAACCTGCCACTCCTGCCCCCTTAACAATCTTGAAAGCCCAACGAGC 749
QY 1937 TCTATGGCAAGAGCTTGTATTCATCCATCGGSCCTGGGAGTAATAACATCAGGACCACT 1996
Db 750 CTTATGGTGTCCAGCCCTGTGTGCCCTGTGGTCCAGGACCAAGAACCAAGATCCACT 809
QY 1997 CGTTTGTATAGTCACTGCTTTTCTTACCATGAAGAAAGAAATCAGATTTTGGCACTATG 2056
Db 810 CTTGTGTACAAACGATTCACCTTTCTCAGCAACACTCCGACCAAGGACTTTTCAACTACA 869
QY 2057 ACTTTAGCAACCTCAGCAGTGTGGCTCATTAATGAATGGCCCGCAGCTTCACCTCCAAG 2116
Db 870 ACTTCTCCGCTTGGCAACACTGTCACTCTGTCTGGAGGCCAAGCTTCACCTCCAAG 929
QY 2117 GAACAAATACCTTCATTTCTCAATATCAGTTTATGTGGCATGAGGGGAAGAAGATGG 2176
Db 930 GGCTGAATATCTCCATCACTTTACCTCAGTCTGTGTGGAACCCAGGCTAGGAAATGT 989
QY 2177 CTTCTGTACCAACATATACAGACTTTACAGTAAAGAAATAGTGGCAGGTCAGATG 2236
Db 990 CTTGTGTGACCAACATGTCACTGACCTCCGGAATTCCTGAGGGTCACTCAGGG----- 1042
QY 2237 ATTACACAAATTTGGTAGGGCATTTGTATGCCAGTCAACAAATTTCTCTCTGAAAGTA 2296
Db 1043 --TTCTCCAAATCTATCAGAGCTACGTCTGCCAGGCACTATCATCCCCCAGAGGTGA 1100
QY 2297 AGGTTTCCGAGACGCTTATCATCAATCATCATTTCTGTGCGAGATCATTTATAGGAG 2356
Db 1101 CAGGCTACAAGCCGGGTTTCTTCCACAGCTGTGAGGCTTCTGATCAGCTTATTTGGG 1160
QY 2357 TCACAGTTGAACACCATGTAAGAAATATTAATATAAAGAAATATGTTCACAGTTCGA 2416
Db 1161 TGACACAGATATGACTCTGGATGAACCTCCCTCCCAAGTAACTTTTCCACCTGGAGT 1220
QY 2417 CAAGCAATATACAGATGTGCACTTTCTTTATAAGTCTTCTACAGCAACAACTTTGTA 2476
Db 1221 CCTTGGGAATACCGACGCTGATCTTTTATAGTCCANTGATGTGACCCAGTCTCTGA 1280
QY 2477 TTAATGGCGATCACTGCTGTGAAATAGAGTGTAACTCTACTAACTTGGAGCAGGAG 2536
Db 1281 GTTCTGGAGATCAACCACTCCGCTCAGGTGAGTCCACAGAAACTGTCCCTGGAA 1340
QY 2537 TGATTTTCACTCCCAAGTCCCAAGCAGGTACCTGTGATGGGTGTACGTTCTATTTC 2596
Db 1341 GTTGTGCTGCCGAAGAGTCTCAGATGGGACCTGTGATGGCTGCAACTTCCACTTCC 1400
QY 2597 TGTGGAGAGTCTGAAGCTTCCCTCTGTGTACGAGCATCTTCCATGAGATTGAGG 2656
Db 1401 TGTGGAGAGCGGCTGCTTCCCGCTCTGTCTAGTGGCTGACTACCATGATATCTGCA 1460
QY 2657 GAGCCTGCAAGAGAGATTTTCAAGAACTTGTATGTGTGGAAATCAACCTAAATGGTGA 2716
Db 1461 GCAGCTGTGTGGCTGATCCAGAGACTTACTTACGTGTGGCGAGAACCCAGCTATGCT 1520
QY 2717 TTAAGGAATTTCTTGGCTGAGAAAGTTGGCAACCTGTGAAACGGTTGACTTTTGGC 2776
Db 1521 CTTGGTGGCAATTTCTTGGCTGAGAGAGAGTCAACCTATGCAAAACCACTAGATTTCTGGC 1580
QY 2777 TGAAGTGGGAGCGGCTGTGGGAGCTTTTACTTGGCGTTTGTGGTGTGGCTCTGACCTGTCT 2836
Db 1581 TGAAGTGGGCACTCTCGAGGCACTGTACTTGCCATCTCTGCTCACCCTCTTGACCTGCT 1640
QY 2837 ACTTGTGAAAAAGATCAAAAGAAAGAGACCACTTTTGAATCTGTTC 2887
Db 1641 ACTTTTGGAAAAAGATCAAAAGTACAGTACCAAGTCTGCAAGCTGGTGA 1691

RESULT 12
AC002081/c

LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AC002081 136150 bp DNA linear PRI 21-DEC-1999
Homo sapiens BAC clone CTA-331C24 from 7q21, complete sequence.
AC002081
AC002081.1 GI:2078453
HTG.
Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 136150)
Du,Z. and Maggii,L.
The sequence of Homo sapiens BAC clone CTA-331C24
Unpublished
2 (bases 1 to 136150)
Waterston,R.
Direct Submission
Submitted (12-MAY-1997) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
3 (bases 1 to 136150)
Waterston,R.
Direct Submission
Submitted (24-SEP-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
4 (bases 1 to 136150)
Waterston,R.
Direct Submission
Submitted (21-DEC-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@watson.wustl.edu
----- Summary Statistics

Center project name: H_RG331C24

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/OTB/CHR7> or <mailto:regreen@nhgri.nih.gov>

SOURCE INFORMATION:

Clone CTA-331C24 is from a release of the human BAC library CTRB-HS-A. The library contains cloned DNA from human sperm. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (<http://www.resgen.com>).
VECTOR: pBelBAC11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is CTA-369H2, 200 bp overlap.
Actual start of this clone is at base position 1 of CTA-331C24;
actual end is at 136150 of CTA-331C24.

Location/Qualifiers
1..136150
/organism="Homo sapiens"

FEATURES
source


```
|||||
Db 396 CCATATTTTAGGCTGTGCTCCTCATACCTGTGCATTTGGTGATCTCACAGAGAGGGCCAT 337
QY 3039 GCCGCTGAAAGGGAAGGAGATTGAAACATTTGATTGGCTTATCACATGGTCAAGTACCT-3098
Db 336 GCCGCTGAAAGGGAAGGAGATTGAAACATTTGATTGGCTTATCACATGGTCAAGTACCT 277
QY 3099 TCCCAATTAAGGGAAGCAAAATGATTGGGCTCAACTGAAGATGAAGCTCAACTCAGGA 3158
Db 276 TGCCAATTAAGGGAAGCAAAATGATTGGGCTCAACTGAAGATGAAGCTCAACTCAGGA 217
QY 3159 AGAGATTTTATCTGTATATACACATAACTGAAACCAAGTTTAAAGCCACCAATGCACCTGC 3218
Db 216 AGAGATTTTATCTGTATATACACATAACTGAAACCAAGTTTAAAGCCACCAATGCACCTGC 157
QY 3219 TGATGCATGCCATATAAATTAATGGGTAACCTTTTATCTTTATGATGTCTACATAACAAGT 3278
Db 156 TGATGCATGCCATATAAATTAATGGGTAACCTTTTATCTTTATGATGTCTACATAACAAGT 97
QY 3279 GTGATTTGGAAGGCACATGTGAGCATATGCATTATGATCCAATTTATGTTTTTCTTTTGT 3338
Db 96 GTGATTTGGAAGGCACATGTGAGCATATGCATTATGATCCAATTTATGTTTTTCTTTTGT 37
QY 3339 TTATATTTTGGGAAAATTAATAATTTTTTAAAGGTA 3374
Db 36 TTATATTTTGGGAAAATTAATAATTTTTTAAAGGTA 1
```

Search completed: May 12, 2003, 05:30:27
Job time : 8660.83 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 23:19:12 ; Search time 674.486 seconds
(without alignments)
11318.654 Million cell updates/sec

Title: US-10-073-333A-1

Perfect score: 3390

Sequence: 1 atgtgttcgcgcgcgggg.....ggtaaaaaaaaaaaaaaaaa 3390

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3390	100.0	3390	22 AAF75054	TR16-short recepto
2	3214	94.8	3556	22 AAF75055	TR16-long receptor
3	1499.6	44.2	1737	22 AAC84891	Human SEC10 nucleic
4	1182.6	34.9	1508	22 AAC84886	Human SEC5 nucleic
5	904.6	26.7	3280	21 AAF95442	Human CASB619 prot
6	903	26.6	3331	21 AAF89765	Nucleotide sequenc
7	901.4	26.6	3334	22 AAF28030	Human TR13 recepto
8	901	26.6	3501	22 AAF21262	Human cDNA sequenc
9	860.8	25.4	2610	22 AAF89775	Nucleotide sequenc

10	860.8	25.4	2733	22 AAF89774	Nucleotide sequenc
11	686.6	20.3	2554	22 AAF27997	Human TR13 recepto
12	493.2	14.5	1129	20 AAX27340	Human secreted pro
13	482.2	14.5	672	22 AAS34165	Human cDNA encodin
14	482.2	14.2	576	24 ABL69028	Kidney cancer rela
15	400.2	11.8	1717	20 AAZ41991	Human endometrium
16	381.8	11.3	1149	22 AAF89777	Nucleotide sequenc
17	379.4	11.2	1299	23 ABV23150	Human prostate exp
18	379.4	11.2	1299	23 ABV28991	Human prostate exp
19	285.8	8.4	1119	22 AAH76196	Human seven-transm
20	285.8	8.4	1587	22 AAH76195	Human seven-transm
21	244	7.2	244	22 ABA71400	Human foetal liver
22	244	7.2	244	22 ABA37625	Probe #16091 for g
23	244	7.2	244	22 AAK19720	Human brain expres
24	244	7.2	244	22 AAK5740	Human bone marrow
25	244	7.2	244	22 AAI51665	Probe #20351 used
26	244	7.2	244	24 ABS20017	Human genome-deriv
27	234.6	6.9	627	22 AAF89776	Nucleotide sequenc
28	218	6.4	371	22 ABA58868	Human foetal liver
29	218	6.4	371	22 ABA27759	Probe #6225 for ge
30	218	6.4	371	22 AAK07028	Human brain expres
31	218	6.4	371	22 AAK32769	Human bone marrow
32	218	6.4	371	22 AAI38582	Probe #7268 used t
33	218	6.4	371	24 ABS07567	Novel human polynu
34	198	5.8	404	22 AAF66437	Human bone marrow
35	196.4	5.8	457	22 AAK32324	Human bone marrow
36	196.4	5.8	457	24 ABS07104	Human genome-deriv
37	196	5.8	196	22 AAK45376	Human bone marrow
38	182.8	5.4	196	24 ABS19637	Human genome-deriv
39	182.8	5.4	210	22 ABA70282	Human foetal liver
40	182.8	5.4	210	22 ABA36993	Probe #15459 for g
41	182.8	5.4	210	22 AAK18511	Human brain expres
42	182.8	5.4	210	22 AAK44426	Human bone marrow
43	182.8	5.4	210	22 AAI50420	Probe #19106 used
44	182.8	5.4	210	24 ABS18661	Human genome-deriv
45	177.8	5.2	457	22 ABA57660	Human foetal liver

ALIGNMENTS

RESULT 1

AAF75054

ID AAF75054 standard; DNA; 3390 BP.

XX AAF75054;

AC AAF75054;

XX

DT 10-MAY-2001 (first entry)

XX

DE TR16-short receptor DNA.

XX

DE

XX

XX

KW TR16 receptor: tumour necrosis factor receptor superfamily;

KW apoptosis; inflammatory; cancer; immune; neurodegenerative; ds.

XX

OS Unidentified.

XX

XX

PN WO200112671-A1.

XX

PD 22-FEB-2001.

XX

XX

PF 10-AUG-2000; 2000WO-US21885.

XX

PR 12-AUG-1999; 99US-0148348.

PR 13-AUG-1999; 99US-0148683.

PR 13-AUG-1999; 99US-0148870.

PR 16-AUG-1999; 99US-0148758.

PR 17-AUG-1999; 99US-0149181.

PR 18-AUG-1999; 99US-0149453.

PR 19-AUG-1999; 99US-0149498.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

XX Ruben SM, Young PE, Baker KP;

PI

XX WPI: 2001-138754/14.
XX New nucleic acid molecule encoding a TR16 tumor necrosis factor
PT receptor polypeptide, useful for the diagnosis and treatment of cancer,
PT autoimmune disorders and cardiovascular diseases -
XX
PS Disclosure: Fig 1: 286pp: English.

XX The present invention relates to a TR16 receptor (tumor necrosis
CC factor receptor superfamily). The invention is useful treating
CC diseases and disorders associated with the inhibited or increased
CC apoptosis. In particular inflammatory diseases, cancers, immune and
CC neurodegenerative disorders may be treated.

SQ Sequence 3390 BP; 979 A; 722 C; 801 G; 888 T; 0 other;

Query Match 100.0%; Score 3390; DB 22; Length 3390;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3390; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCTGTTCCGCGCGCGCGCGGTACGCGGCGGCGCTGGGGCGCGCGGAGGCT 60
DB 1 ATGCTGTTCCGCGCGCGCGCGGTACGCGGCGGCGCTGGGGCGCGCGGAGGCT 60
QY 61 CCCGCGCGCGCGCTGCGCCCGCTGGAGCCCGCGCTGGATTGCTGCGCGCTCGCC 120
DB 61 CCCGCGCGCGCGCTGCGCCCGCTGGAGCCCGCGCTGGATTGCTGCGCGCTCGCC 120
QY 121 GCGTCCGCGCGCGCTGGGCTGGGACCTGCGCTCTCTCCAGCGCGCGCTCTCTCT 180
DB 121 GCGTCCGCGCGCGCTGGGCTGGGACCTGCGCTCTCTCCAGCGCGCGCTCTCTCT 180
QY 181 TGCCAGGAGAAAGATTATCACATTGAATACGGAATGATAGCAGTGGCTCCAGGTGG 240
DB 181 TGCCAGGAGAAAGATTATCACATTGAATACGGAATGATAGCAGTGGCTCCAGGTGG 240
QY 241 AGAGTTGCCATTCCAAATTCAGTGGAGTCTCTGGGCTCGCTGACCCAGTGAGAGGC 300
DB 241 AGAGTTGCCATTCCAAATTCAGTGGAGTCTCTGGGCTCGCTGACCCAGTGAGAGGC 300
QY 301 AAAGAAATGACATTTCTCTCTGCTGCGAGAGTATCTAGAAATGAAGAACAGGTATGC 360
DB 301 AAAGAAATGACATTTCTCTCTGCTGCGAGAGTATCTAGAAATGAAGAACAGGTATGC 360
QY 361 AGTAAGTGGTGAAGGCACCTATTCTTGGCAGTGGCATCAAAATTTGATGAATGGAT 420
DB 361 AGTAAGTGGTGAAGGCACCTATTCTTGGCAGTGGCATCAAAATTTGATGAATGGAT 420
QY 421 GAATTCGCGCGAGGATTTCTAACATCGCAACATTCATGCGACACTGTGTGGGCCCTTCT 480
DB 421 GAATTCGCGCGAGGATTTCTAACATCGCAACATTCATGCGACACTGTGTGGGCCCTTCT 480
QY 481 GACAGCGCCGACAGCGCTGTAACAACTCTCTTGGATCCCTCGTGGAACTACATAGAA 540
DB 481 GACAGCGCCGACAGCGCTGTAACAACTCTCTTGGATCCCTCGTGGAACTACATAGAA 540
QY 541 TCTAATCGTGATGACCTGACCGTGTCTTGTATCTATGCTGCACCTTAAGAAGTCAGCG 600
DB 541 TCTAATCGTGATGACCTGACCGTGTCTTGTATCTATGCTGCACCTTAAGAAGTCAGCG 600
QY 601 TATGCTCTCTTTGAGTACAGTATGTCGACACACACATCTCTTTGAGTCTTTATTCAA 560
DB 601 TATGCTCTCTTTGAGTACAGTATGTCGACACACACATCTCTTTGAGTCTTTATTCAA 560
QY 661 AATGATCAGTCCGAGAGATGGACACCACTGACAACTGGGTAAACCTTACAGACAT 720
DB 661 AATGATCAGTCCGAGAGATGGACACCACTGACAACTGGGTAAACCTTACAGACAT 720
QY 721 GGAGAAATGGGCTCTCATTTCTGTAATGCTGAATCAGGCACAAACATCTACTTGAGA 780
DB 721 GGAGAAATGGGCTCTCATTTCTGTAATGCTGAATCAGGCACAAACATCTACTTGAGA 780

QY 781 ACTACAGGCATCTTATGGGTTCTAAGCGGTCAAGCCTGTGCTGTGTAATAAATATACA 840
DB 781 ACTACAGGCATCTTATGGGTTCTAAGCGGTCAAGCCTGTGCTGTGTAATAAATATACA 840
QY 841 ATTGAAGGGTGGCGTACACATCAGAAATGTTTCTTGAAGCCAGGCACATTCAGCAAC 900
DB 841 ATTGAAGGGTGGCGTACACATCAGAAATGTTTCTTGAAGCCAGGCACATTCAGCAAC 900
QY 901 AAACAGGTTTCAATCAACCTCCAGGTGTCTCCAGAACACACCTATTTCTCAGAAAGAGCC 960
DB 901 AAACAGGTTTCAATCAACCTCCAGGTGTCTCCAGAACACACCTATTTCTCAGAAAGAGCC 960
QY 961 AAAGAATGTATGAAGGTGAAGACGACTCTCAATTTTTCAGGATCCAGTGAGTACAGAG 1020
DB 961 AAAGAATGTATGAAGGTGAAGACGACTCTCAATTTTTCAGGATCCAGTGAGTACAGAG 1020
QY 1021 CCGCCTCCCTGTACCACAAAGACTATTTCCAGATCCATCTCCATGTGATCAAGAAGCA 1080
DB 1021 CCGCCTCCCTGTACCACAAAGACTATTTCCAGATCCATCTCCATGTGATCAAGAAGCA 1080
QY 1081 AAGACACAGATAATGTACAAGTGGATAGAGCCAAATCTGCCGGGAGGATCTCACAGAT 1140
DB 1081 AAGACACAGATAATGTACAAGTGGATAGAGCCAAATCTGCCGGGAGGATCTCACAGAT 1140
QY 1141 GCTATTAGATGCCCGCTTCTCGAGAGAAAGGATTTGCCGCTTGGCAACCTGGATT 1200
DB 1141 GCTATTAGATGCCCGCTTCTCGAGAGAAAGGATTTGCCGCTTGGCAACCTGGATT 1200
QY 1201 TATACAATGGATCATCTTCTTGGCATCCCTGCTCTCTCGAACAATTTTCAGATGGAACC 1260
DB 1201 TATACAATGGATCATCTTCTTGGCATCCCTGCTCTCTCGAACAATTTTCAGATGGAACC 1260
QY 1261 AAAGAATGTAGACCATGTCCAGCAGGAGGAGCCTGCATCTTGGCTTTGAATATAAATGG 1320
DB 1261 AAAGAATGTAGACCATGTCCAGCAGGAGGAGCCTGCATCTTGGCTTTGAATATAAATGG 1320
QY 1321 TGAATGTCCTTCTCGGCAACATGAAACCTTCTGCTCAATGTTGGGAATTCAAAGTGC 1380
DB 1321 TGAATGTCCTTCTCGGCAACATGAAACCTTCTGCTCAATGTTGGGAATTCAAAGTGC 1380
QY 1381 GATGAATCAATGGTGGGAGTGGCTGGAGATCATATCCAGATGGGCTGGAGTTCT 1440
DB 1381 GATGAATCAATGGTGGGAGTGGCTGGAGATCATATCCAGATGGGCTGGAGTTCT 1440
QY 1441 GACAATGATTACCTGATCTTAACTTTGCATATCCAGAGATTTAAACCCACCAATCTATG 1500
DB 1441 GACAATGATTACCTGATCTTAACTTTGCATATCCAGAGATTTAAACCCACCAATCTATG 1500
QY 1501 ACTGGAGCCAGGGTCTCAACTAGGAAGATAACATTTGCTTTTGGAGACCTCTCTTCA 1560
DB 1501 ACTGGAGCCAGGGTCTCAACTAGGAAGATAACATTTGCTTTTGGAGACCTCTCTTCA 1560
QY 1561 GCTGACTGTGTTTGTACTTCAATGGTGGATATTAAGAAAAGTACAATGTGGTAGAA 1620
DB 1561 GCTGACTGTGTTTGTACTTCAATGGTGGATATTAAGAAAAGTACAATGTGGTAGAA 1620
QY 1621 TCGTGGGTGGAACCAAGAAACAAAGCTTACCCCATATCATCTTCAAGAAATGCAACT 1680
DB 1621 TCGTGGGTGGAACCAAGAAACAAAGCTTACCCCATATCATCTTCAAGAAATGCAACT 1680
QY 1681 TTTTACATTTACATGGGCATTCAGAGAACTAATCAGGGTCAAGATTAATAGACGGTTTCATC 1740
DB 1681 TTTTACATTTACATGGGCATTCAGAGAACTAATCAGGGTCAAGATTAATAGACGGTTTCATC 1740
QY 1741 AATGACATGGTGAAGATTTATTTCTATCAGAGCACTTAATGCAGTTGATGGGTGGCGTCC 1800
DB 1741 AATGACATGGTGAAGATTTATTTCTATCAGAGCACTTAATGCAGTTGATGGGTGGCGTCC 1800
QY 1801 TCATGCCGTGCGCTGCGCTCGGTTCTGAACAGTGGGTTTCATGCGTGTGTCGCCCTGCCCT 1860
DB 1801 TCATGCCGTGCGCTGCGCTCGGTTCTGAACAGTGGGTTTCATGCGTGTGTCGCCCTGCCCT 1860
QY 1861 CCAGGCCACTACATTTGAGAAAGAAACCAACCACTGCAAGNAATGTCCACCTGACACCTAC 1920

|||||
Db 1861 CCAGGCCACTACATTGAGAAAGAAACCAACAGTGCAGGAATGTCCACTGACACCTAC 1920
QY 1921 CTGTCCATACATCAGTCTATGGCAAGAGGCTGTATTCCATGGGGCTGGGAGTAA 1980
Db 1921 CTGTCCATACATCAGTCTATGGCAAGAGGCTGTATTCCATGGGGCTGGGAGTAA 1980
QY 1981 AACAAATCAGGACCAATTCGGTTTCTATAGTACTGCTTTTCTACCATGAAAGAAAT 2040
Db 1981 AACAAATCAGGACCAATTCGGTTTCTATAGTACTGCTTTTCTACCATGAAAGAAAT 2040
QY 2041 CAGATTTTGCATATGACATTTAGCAACCTCAGAGTGGGGCTCATTAATGAATGCCCC 2100
Db 2041 CAGATTTTGCATATGACATTTAGCAACCTCAGAGTGGGGCTCATTAATGAATGCCCC 2100
QY 2101 AGCTTCACCTCCAAAGGAACAAATACTTCCATTTCTCAATATCATAGTTTATGTGGCAT 2160
Db 2101 AGCTTCACCTCCAAAGGAACAAATACTTCCATTTCTCAATATCATAGTTTATGTGGCAT 2160
QY 2161 GAGGGGAAGAAGTGGCTCTCTGTACCAACAATATAACAGACTTTACAGTAAAGAAATA 2220
Db 2161 GAGGGGAAGAAGTGGCTCTCTGTACCAACAATATAACAGACTTTACAGTAAAGAAATA 2220
QY 2221 GTGGCAGGTCAGATGATTACACAAATTTGGTAGGGGCAATTTGTATGCCAGTCAACAAT 2280
Db 2221 GTGGCAGGTCAGATGATTACACAAATTTGGTAGGGGCAATTTGTATGCCAGTCAACAAT 2280
QY 2281 ATTCTCTCTGAAAGTAAAGGTTTCCAGCAGCTTATCATCAATCCATCATCTGGCA 2340
Db 2281 ATTCTCTCTGAAAGTAAAGGTTTCCAGCAGCTTATCATCAATCCATCATCTGGCA 2340
QY 2341 GATACATTCATAGGAGTCAGATTGAAACCAATTAACAGACTTTACAGTAAAGAAAT 2400
Db 2341 GATACATTCATAGGAGTCAGATTGAAACCAATTAACAGACTTTACAGTAAAGAAAT 2400
QY 2401 ATCTTCCAGTTCCACAGCCAAATACAGATGTCATTTCTTTATAGTCTTCTACA 2460
Db 2401 ATCTTCCAGTTCCACAGCCAAATACAGATGTCATTTCTTTATAGTCTTCTACA 2460
QY 2461 GCAACAACATCTGTAATAGTCCCATCACTGCTGCAAAATGAGTGTATCTACT 2520
Db 2461 GCAACAACATCTGTAATAGTCCCATCACTGCTGCAAAATGAGTGTATCTACT 2520
QY 2521 AAATCGAGCAGGAGTATTTAGTCCCAAGTGCAGCAGAGTACCTGTGATGG 2580
Db 2521 AAATCGAGCAGGAGTATTTAGTCCCAAGTGCAGCAGAGTACCTGTGATGG 2580
QY 2581 TGTACGTTCTATTTCTGTGGGAGTGTGAAGCTTGGCCCTCTGTGTACGGAGCATGAC 2640
Db 2581 TGTACGTTCTATTTCTGTGGGAGTGTGAAGCTTGGCCCTCTGTGTACGGAGCATGAC 2640
QY 2641 TTCCATGAGATTGAGGAGCCTGCAAGAGAGATTTCAGGAAACCTTGTATGTGGAAAT 2700
Db 2641 TTCCATGAGATTGAGGAGCCTGCAAGAGAGATTTCAGGAAACCTTGTATGTGGAAAT 2700
QY 2701 GAACCTAAATGTTGCAATTAAGAAATTTCTTTGCCGTGAGAAAAGTTGCCAACCTGTGAA 2760
Db 2701 GAACCTAAATGTTGCAATTAAGAAATTTCTTTGCCGTGAGAAAAGTTGCCAACCTGTGAA 2760
QY 2761 ACGGTTGACTTTTGGCTGAAGTGGGAGCCGTTGGGGAGCTTTTACTCCGTTTGTG 2820
Db 2761 ACGGTTGACTTTTGGCTGAAGTGGGAGCCGTTGGGGAGCTTTTACTCCGTTTGTG 2820
QY 2821 GTGGCTCTGACCTGTACTTCTGGAAAAAGAAATCAAAAGAAAGAACCATTTTGAAT 2880
Db 2821 GTGGCTCTGACCTGTACTTCTGGAAAAAGAAATCAAAAGAAAGAACCATTTTGAAT 2880
QY 2881 CTGTTCAACTGAAACCTCAAGATCCCCAAATATATGAAGACAGTGTGAGCCTTGA 2940
Db 2881 CTGTTCAACTGAAACCTCAAGATCCCCAAATATATGAAGACAGTGTGAGCCTTGA 2940
QY 2941 GACTATGAACAAAGAACTGCTGTAGTTTACAGGACCATATTTAGGGTCTGTCTTC 3000
|||||

Db 2941 GACTAATGAACAAAGAAACCTGCTGTAGTTTACAGGACCATATTTAGGGTCTGTCTC 3000
QY 3001 ATACCTGTACACATTGGTGATCTCACAGAGAGGGCCATCCGCTGAAAAGGAGAGAT 3060
Db 3001 ATACCTGTACACATTGGTGATCTCACAGAGAGGGCCATCCGCTGAAAAGGAGAGAT 3060
QY 3061 TGAACATATTGATTGCCCTTATCACATGGTCAAGTACCTTGCACAAATAAAGGAAACAAAT 3120
Db 3061 TGAACATATTGATTGCCCTTATCACATGGTCAAGTACCTTGCACAAATAAAGGAAACAAAT 3120
QY 3121 GATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAGGAAGAGATTTATCTGTATATAC 3180
Db 3121 GATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAGGAAGAGATTTATCTGTATATAC 3180
QY 3181 ATAACTGAAAAACCAAGTTTAAAGCCCAACCAATGCACCTGCTGATGCATGCCATATAAT 3240
Db 3181 ATAACTGAAAAACCAAGTTTAAAGCCCAACCAATGCACCTGCTGATGCATGCCATATAAT 3240
QY 3241 GGGTAACCTTTTATCTTTATGATGTCTACATAACAAGTGTGATTTGGAAGGCACATGTA 3300
Db 3241 GGGTAACCTTTTATCTTTATGATGTCTACATAACAAGTGTGATTTGGAAGGCACATGTA 3300
QY 3301 GCATATGCATTTATGATCCAAATTTATGTTTTTCTTTTGTATATTTTGGGAAAAATTA 3360
Db 3301 GCATATGCATTTATGATCCAAATTTATGTTTTTCTTTTGTATATTTTGGGAAAAATTA 3360
QY 3361 ATTTTAAAGTAAAAAAGAAAAA 3390
Db 3361 ATTTTAAAGTAAAAAAGAAAAA 3390

RESULT 2
AAF75055
ID AAF75055 standard; DNA; 3556 BP.
XX AAF75055;
AC AAF75055;
DT 10-MAY-2001 (first entry)
XX TR16-long receptor DNA.
DE TR16 receptor; tumour necrosis factor receptor superfamily;
KW apoptosis; inflammatory; cancer; immune; neurodegenerative; ds.
XX Unidentified.
OS
PN WO200112671-A1.
XX
PD 22-FEB-2001.
XX
PF 10-AUG-2000; 2000WO-US21885.
XX
PR 12-AUG-1999; 99US-0148348.
PR 13-AUG-1999; 99US-0148683.
PR 13-AUG-1999; 99US-0148870.
PR 16-AUG-1999; 99US-0148758.
PR 17-AUG-1999; 99US-0149181.
PR 18-AUG-1999; 99US-0149453.
PR 19-AUG-1999; 99US-0149498.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Young PE, Baker KP;
XX WPI; 2001-138754/14.
XX
PT New nucleic acid molecule encoding a TR16 tumor necrosis factor
PT receptor polypeptide, useful for the diagnosis and treatment of cancer,
PT autoimmune disorders and cardiovascular diseases
XX
PS Disclosure: Fig 4; 286pp; English.
XX
CC The present invention relates to a TR16 receptor (tumour necrosis

CC factor receptor superfamily). The invention is useful treating
CC diseases and disorders associated with the inhibited or increased
CC apoptosis. In particular inflammatory diseases, cancers, immune and
XX neurodegenerative disorders may be treated.

..SQ Sequence 3556 BP: 1043 A; 751 C; 836 G; 926 T; 0 other;

Query Match 94.8%; Score 3214; DB 22; Length 3556;
Best Local Similarity 95.3%; Pred. No. 0;
Matches 3390; Conservative 0; Mismatches 0; Indels 166; Gaps 1;

```
QY 1 ATGCTGTTCCGCGCGCGCGCGGTAGGGGACAGGGCTGGGGCGCGCGGAGGCT 60
DQ 1 ATGCTGTTCCGCGCGCGCGCGGTAGGGGACAGGGCTGGGGCGCGCGGAGGCT 60
QY 61 CCCCGCGCGCGCGCTCGCGCCCTGGAGCCCGCCCTGGAATTTGCTGGCGCTCGCC 120
DQ 61 CCCCGCGCGCGCGCTCGCGCCCTGGAGCCCGCCCTGGAATTTGCTGGCGCTCGCC 120
QY 121 GGCTGCCAGCGCGCTGGGCTGGGACCTGCCCTCCTCCTCCAGCGCGCGCTTCTCT 180
DQ 121 GGCTGCCAGCGCGCTGGGCTGGGACCTGCCCTCCTCCTCCAGCGCGCGCTTCTCT 180
QY 181 TGCCAGGAGAAAGATATACATTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGG 240
DQ 181 TGCCAGGAGAAAGATATACATTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGG 240
QY 241 AGAGTTGCCATTCCAAATTCGACGTGGAAGCTCTGGCTGCCCTGACCCAGTGAGAGGC 300
DQ 241 AGAGTTGCCATTCCAAATTCGACGTGGAAGCTCTGGCTGCCCTGACCCAGTGAGAGGC 300
QY 301 AAAGAATGCACATTTCTCCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAACAGGTATGC 360
DQ 301 AAAGAATGCACATTTCTCCTGTGCTTCTGGAGAGTATCTAGAAATGAAGAACAGGTATGC 360
QY 361 AGTAAGTGTGTGAGGACCATTTCTTGGCGAGTGGCAGTGGCATCAAAATTTGATGAGGAT 420
DQ 361 AGTAAGTGTGTGAGGACCATTTCTTGGCGAGTGGCAGTGGCATCAAAATTTGATGAGGAT 420
QY 421 GAATTTGCCGCGAGGATTTTCTAACATCGCAACATTCATGGACACTGTGGCGCCCTTCT 480
DQ 421 GAATTTGCCGCGAGGATTTTCTAACATCGCAACATTCATGGACACTGTGGCGCCCTTCT 480
QY 481 GACAGCGCCAGAGCGGTGTAAACACTCTTCTGGATCCCTCGTGGAACTACATAGAA 540
DQ 481 GACAGCGCCAGAGCGGTGTAAACACTCTTCTGGATCCCTCGTGGAACTACATAGAA 540
QY 541 TCTAATCGTGATGACTGCACGCTGTCTTGTATCTATGCTGCACCTTAAGAAGTCAGGC 600
DQ 541 TCTAATCGTGATGACTGCACGCTGTCTTGTATCTATGCTGCACCTTAAGAAGTCAGGC 600
QY 601 TATGCTCTTTGAGTACAGTATGTGACAAACATCTTCTTGTGAGTTCTTTATTCAA 660
DQ 601 TATGCTCTTTGAGTACAGTATGTGACAAACATCTTCTTGTGAGTTCTTTATTCAA 660
QY 661 AATGATCAGTGCAGAGATGACACACCACTGACAAAGTGGGTAAACTTACAGACAA 720
DQ 661 AATGATCAGTGCAGAGATGACACACCACTGACAAAGTGGGTAAACTTACAGACAA 720
QY 721 GCAGAAATGGGCTCTATCTCTAATGCTGAATCAGGCACAAACATCTCTACTGGAGA 780
DQ 721 GCAGAAATGGGCTCTATCTCTAATGCTGAATCAGGCACAAACATCTCTACTGGAGA 780
QY 781 ACTACAGGCATCTTATGGGTTCTAAGCGGTCAAGCGCTGTGCTGGTAAAAATATCA 840
DQ 781 ACTACAGGCATCTTATGGGTTCTAAGCGGTCAAGCGCTGTGCTGGTAAAAATATCA 840
QY 841 ATTGAAGGGGTGGGTACACATACAGAAATGTTTCTTGAAGCGGACAGGACATTCAGCAAC 900
DQ 841 ATTGAAGGGGTGGGTACACATACAGAAATGTTTCTTGAAGCGGACAGGACATTCAGCAAC 900
QY 901 AAACAGGTTTCATTCACCTGCCAGGTGTCTCCAGAAACACCATTTCTGAGAAAGGAGCC 960
DQ 901 AAACAGGTTTCATTCACCTGCCAGGTGTCTCCAGAAACACCATTTCTGAGAAAGGAGCC 960
```

```
Db 901 AAACAGGTTTCATTCACCTGCCAGGTGTCTCCAGAAACACCATTTCTGAGAAAGGAGCC 960
QY 961 AAAGAATGTATAAGGTGTAAAGACGACTCTCAATTTTTCAGGATCCAGTGTACAGAG 1020
DQ 961 AAAGAATGTATAAGGTGTAAAGACGACTCTCAATTTTTCAGGATCCAGTGTACAGAG 1020
QY 1021 CGCCCTCCCTGTACACAAAGACTATTTCAGATCCATCTACTCCATGTGATGAAGAAGGA 1080
DQ 1021 CGCCCTCCCTGTACACAAAGACTATTTCAGATCCATCTACTCCATGTGATGAAGAAGGA 1080
QY 1081 AAGACACAGATTAATGTACAAAGTGGATAGAGCCCAAAATCTGCGGGAGGATCTCACAGAT 1140
DQ 1081 AAGACACAGATTAATGTACAAAGTGGATAGAGCCCAAAATCTGCGGGAGGATCTCACAGAT 1140
QY 1141 GCTATTAGATTGCCCCCTTCTGGAGAGAAAGAAATGTGTCGCCCTTGAACCCCTCGATT 1200
DQ 1141 GCTATTAGATTGCCCCCTTCTGGAGAGAAAGAAATGTGTCGCCCTTGAACCCCTCGATT 1200
QY 1201 TATAACAATGGATCATCTTCTGCCATCCCTGTCTCTCGTGGAAACATTTTCAGATGGAACC 1260
DQ 1201 TATAACAATGGATCATCTTCTGCCATCCCTGTCTCTCGTGGAAACATTTTCAGATGGAACC 1260
QY 1261 AAAGAATGTAGACCATGTCCAGCAGGAAAGAGCTGACATTTGGCTTTGAATATAAATGG 1320
DQ 1261 AAAGAATGTAGACCATGTCCAGCAGGAAAGAGCTGACATTTGGCTTTGAATATAAATGG 1320
QY 1321 TGGAAATGTCCTTCCCTGGCAACATGAAATCTCTCTCAATTTGGGAATTTCAAAGTGC 1380
DQ 1321 TGGAAATGTCCTTCCCTGGCAACATGAAATCTCTCTCAATTTGGGAATTTCAAAGTGC 1380
QY 1381 GATGAATGAATGTTGGAGGTGGCTGGAGATCATATCCAGAGTGGGCTGGAGTTCT 1440
DQ 1381 GATGAATGAATGTTGGAGGTGGCTGGAGATCATATCCAGAGTGGGCTGGAGTTCT 1440
QY 1441 GACAATGATTAACCTCATCTTAACTTGCATATCCAGGATTTAAACACCAACATCTATG 1500
DQ 1441 GACAATGATTAACCTCATCTTAACTTGCATATCCAGGATTTAAACACCAACATCTATG 1500
QY 1501 ACTGGAGCCACGGTTCCTGAACCTAGGAAAGAAATGATTTGTCTTTGAGACCTCTCTTCA 1560
DQ 1501 ACTGGAGCCACGGTTCCTGAACCTAGGAAAGAAATGATTTGTCTTTGAGACCTCTCTTCA 1560
QY 1561 GCTGACTGTGTTTGTACTTCTATGTTGGATATTAATAGAAAAGTACAATTTGGTAGAA 1620
DQ 1561 GCTGACTGTGTTTGTACTTCTATGTTGGATATTAATAGAAAAGTACAATTTGGTAGAA 1620
QY 1621 TCGTGGGTGGAAACCAAGAAAGAAAGCTTACCCCATATCTTCAAGAAATGCAACT 1680
DQ 1621 TCGTGGGTGGAAACCAAGAAAGAAAGCTTACCCCATATCTTCAAGAAATGCAACT 1680
QY 1681 TTTTACATTTTACATGGGCAITCCAGAGAACTAATCAGGCTCAAGATTAATAGACGGTTCA 1740
DQ 1681 TTTTACATTTTACATGGGCAITCCAGAGAACTAATCAGGCTCAAGATTAATAGACGGTTCA 1740
QY 1741 AATGACATGGTGAAGATTTATTTCTATCAGCCACTAATGCAAGTTGATGGGGTGGGCTCC 1800
DQ 1741 AATGACATGGTGAAGATTTATTTCTATCAGCCACTAATGCAAGTTGATGGGGTGGGCTCC 1800
QY 1801 TCATCCGTCGCTGTGCCCTCGTCTGAACAGTCCGGTTTCTATGCTGCTCCCTGCCCT 1860
DQ 1801 TCATCCGTCGCTGTGCCCTCGTCTGAACAGTCCGGTTTCTATGCTGCTCCCTGCCCT 1860
QY 1861 CCAGGCCACTTACATTTGAGAAAGAAACCAACCAAGTGAAGAAATGTCACACTTACACCTAC 1920
DQ 1861 CCAGGCCACTTACATTTGAGAAAGAAACCAACCAAGTGAAGAAATGTCACACTTACACCTAC 1920
QY 1921 CTGTCCATACATCAGGTCTATGCGGAAGAGGCTTGTATTTCCATGGGGCTGGGAGTAA 1980
DQ 1921 CTGTCCATACATCAGGTCTATGCGGAAGAGGCTTGTATTTCCATGGGGCTGGGAGTAA 1980
QY 1981 AACAATCAGGACCATTTCCGTTTGTGCTATGTCATCTTCTACCATGAAAAAGAAAT 2040
DQ 1981 AACAATCAGGACCATTTCCGTTTGTGCTATGTCATCTTCTACCATGAAAAAGAAAT 2040
```

QY 2041 CAGATTTTGCACCTATGACTTTAGCAACCTCAGCAGTGTGGGCTCATTAATGAATGCCCC 2100
|||||
Db 2041 CAGATTTTGCACCTATGACTTTAGCAACCTCAGCAGTGTGGGCTCATTAATGAATGCCCC 2100
QY 2101 AGCTTCACCTCCAAAGGAACAATACTTCCATTTCTTCAATATCAATAGTTATGTGGGCAT 2160
|||||
Db 2101 AGCTTCACCTCCAAAGGAACAATACTTCCATTTCTTCAATATCAATAGTTATGTGGGCAT 2160
QY 2161 GAGGGGAAGAAGTGGCTCTGTGTACCAACAATATAACAGACTTTACAGTAAAGAAATA 2220
|||||
Db 2161 GAGGGGAAGAAGTGGCTCTGTGTACCAACAATATAACAGACTTTACAGTAAAGAAATA 2220
QY 2221 GTGGCAGGCTCAGATGATTACAAATTTGGTAGGGGCAATTTGTATGCCAGTCAACAAT 2280
|||||
Db 2221 GTGGCAGGCTCAGATGATTACAAATTTGGTAGGGGCAATTTGTATGCCAGTCAACAAT 2280
QY 2281 ATTCTTCTGAAAGTAAGGTTTCCGAGCAGGCTTATCATCAATATCAATATCTGTGGA 2340
|||||
Db 2281 ATTCTTCTGAAAGTAAGGTTTCCGAGCAGGCTTATCATCAATATCAATATCTGTGGA 2340
QY 2341 GATACATTCATAGGAGTCACAGTTGAACCCACATTCGAAATAATTAATAAAGAGAT 2400
|||||
Db 2341 GATACATTCATAGGAGTCACAGTTGAACCCACATTCGAAATAATTAATAAAGAGAT 2400
QY 2401 ATGTTCCCAAGTTCCAAACAGCCAAATFACCAGATGTGCATTTCTTTTATAGTCTTCTACA 2460
|||||
Db 2401 ATGTTCCCAAGTTCCAAACAGCCAAATFACCAGATGTGCATTTCTTTTATAGTCTTCTACA 2460
QY 2461 GCACCAACATCTTGATTAATGCCGATCACTGCTGTGGAATGAGGTGTAATCTACT 2520
|||||
Db 2461 GCACCAACATCTTGATTAATGCCGATCACTGCTGTGGAATGAGGTGTAATCTACT 2520
QY 2521 AAATCTGGAGCAGGAGTGTATTCAGTCCCAAGCAGTCCAGCAGGACCTGCTGATGG 2580
|||||
Db 2521 AAATCTGGAGCAGGAGTGTATTCAGTCCCAAGCAGTCCAGCAGGACCTGCTGATGG 2580
QY 2581 TGTACGTTCTATTTCTGTGGGAGTGTGAAGCTTGGCCCTCTGTGTACGGAGCATGAC 2640
|||||
Db 2581 TGTACGTTCTATTTCTGTGGGAGTGTGAAGCTTGGCCCTCTGTGTACGGAGCATGAC 2640
QY 2641 TTCCATGAGATGAGGAGCCCTGCAAGAGAGATTTTCAGAAACCTTGTATGTGGGAAT 2700
|||||
Db 2641 TTCCATGAGATGAGGAGCCCTGCAAGAGAGATTTTCAGAAACCTTGTATGTGGGAAT 2700
QY 2701 GAACCTAAATGTTGCTATTAAGGAATTTCTTTCCTGAGAAAAGTTGCAACCTGTGA 2760
|||||
Db 2701 GAACCTAAATGTTGCTATTAAGGAATTTCTTTCCTGAGAAAAGTTGCAACCTGTGA 2760
QY 2761 ACGGTTGACTTTTGGCTGAAGTGGAGCCGCTGTGGAGCTTTTACTGCCGTTTGTCTG 2820
|||||
Db 2761 ACGGTTGACTTTTGGCTGAAGTGGAGCCGCTGTGGAGCTTTTACTGCCGTTTGTCTG 2820
QY 2821 GTGGCTCTGACCTGCTACTCTTGGAAGAAAGATCAAAA 2858
|||||
Db 2821 GTGGCTCTGACCTGCTACTCTTGGAAGAAAGATCAAAA 2858
QY 2859 2858
Db 2881 TTAGTAATGACGACTAACTCAAAAGAGTGTGAACCTCCCGCTGCAGACAGTTGTGCTATC 2940
QY 2859 2858
Db 2941 ATGGAAGGAGAGATAATGAAGAGGAAGTTGTATATTCCAATAAAGACACTACTAGGA 3000
QY 2859 -----GAAAAGAGACCATTTTGAATCTGTTCAACTGAAA 2894
|||||
Db 3001 AAACCTAAATCTTTGCAACCAAGGAAAGAGACCATTTTGAATCTGTTCAACTGAAA 3060
QY 2895 ACCTCAAGATCCCCAATATATGAAGACAGTGTGCTGTAGCCTTGAGACTAATGACAAA 2954
|||||
Db 3061 ACCTCAAGATCCCCAATATATGAAGACAGTGTGCTGTAGCCTTGAGACTAATGACAAA 3120

QY 2955 GAAACCTGCTAGTGTATTTTACAGACCATATTTAGGCTGTCTCCTCATACCTGTGCACATT 3014
|||||
Db 3121 GAAACCTGCTAGTGTATTTTACAGACCATATTTAGGCTGTCTCCTCATACCTGTGCACATT 3180
QY 3015 GGTGATCTCACAGAGAGGGCCATGCCGTGAAAAGGAGAGATTGAAACATTTGATT 3074
|||||
Db 3181 GGTGATCTCACAGAGAGGGCCATGCCGTGAAAAGGAGAGATTGAAACATTTGATT 3240
QY 3075 GCCTATACATGGTCAAGTACCTTGCCTGCAAAATAAGGAAGCAAAATGATTGGGTCTCAA 3134
|||||
Db 3241 GCCTATACATGGTCAAGTACCTTGCCTGCAAAATAAGGAAGCAAAATGATTGGGTCTCAA 3300
QY 3135 CTGAAGTGAAGTCAACTCAGAGAGAGATTATCTGTATATACACATAACTGAAAAACCA 3194
|||||
Db 3301 CTGAAGTGAAGTCAACTCAGAGAGAGATTATCTGTATATACACATAACTGAAAAACCA 3360
QY 3195 AGTTAAGCCCACTCACTGCTGATGCATGCCATATAAATTAATGGTAACTTTTATT 3254
|||||
Db 3361 AGTTAAGCCCACTCACTGCTGATGCATGCCATATAAATTAATGGTAACTTTTATT 3420
QY 3255 CTTTATGATCTTACATAACAAGTGTGATTGGAAGGCACATGTGAGCATATGCATTATG 3314
|||||
Db 3421 CTTTATGATCTTACATAACAAGTGTGATTGGAAGGCACATGTGAGCATATGCATTATG 3480
QY 3315 ATCCAAATTTATGTTTTTCTTTTCTTTTATATTTTGGGAAAAATTTAAATTTTAAAGGTA 3374
|||||
Db 3481 ATCCAAATTTATGTTTTTCTTTTCTTTTATATTTTGGGAAAAATTTAAATTTTAAAGGTA 3540
QY 3375 AAAAAAATAAAAAA 3390
|||||
Db 3541 AAAAAAATAAAAAA 3556

RESULT 3
AAC84891
ID AAC84891 standard; cdna; 1737 BP.
XX AAC84891;
XX AC
XX 20-APR-2001 (first entry)
XX Human SEC10 nucleic acid sequence (clone ID 1795045.0.77).
DE SECX; cytostatic; gynecological; gene therapy; screening assay; human;
KW chromosomal mapping; forensic biology; cell proliferation; cancer;
KW cell differentiation; immune associated disorder; gestational disease;
KW SEC10; ss.
XX Homo sapiens.
XX OS
XX Key Location/Qualifiers
FH 296..1690
FT CDS /*tag= a
FT /product= "SEC10"
XX WO200078802-A2.
XX PN
XX 28-DEC-2000.
XX PF 23-JUN-2000; 2000WO-US17328.
XX PR 23-JUN-1999; 99US-0140584.
PR 20-JUL-1999; 99US-0144722.
PR 16-SEP-1999; 99US-0154520.
PR 22-JUN-2000; 2000US-0604286.
XX (CURA-) CURAGEN CORP.
XX Shinketsu RA, Fernandes E, Vernet C, Yang M, Boldog FL;
PI Herrmann JL;
XX WPI; 2001-071385/08.
DR P-PSDB; AAB48377.

XX Polynucleotides encoding SECX proteins useful for treating disease
PT characterized by an aberrant level of cell proliferation and/or
PT differentiation like cancer or immune associated disorders -
XX
PS Claim 3; Fig 10; 132pp; English.
XX
CC The invention relates to human SECX polypeptides and polynucleotides
CC encoding them. The SECX polypeptides can be expressed by standard
CC recombinant methodology. The SECX polypeptides are useful for treating
CC or preventing a SECX-associated disorder. The invention is useful in
CC screening assays: detection assays (e.g. chromosomal mapping, cell and
CC tissue typing, forensic biology); predictive medicine (diagnostic assays,
CC prognostic assays, monitoring clinical trials, and pharmacogenomics); and
CC methods of treatment (e.g. therapeutic and prophylactic), especially
CC disorders characterized by aberrant cell proliferation and/or
CC differentiation like cancer or immune associated disorders or gestational
CC disease. The present sequence represents a SEC10 nucleic acid sequence.
XX
SQ Sequence 1737 BP; 495 A; 381 C; 409 G; 451 T; 1 other:

Query Match 44.2%; Score 1499.6; DB 22; Length 1737;
Best Local Similarity 98.7%; Pred. No. 0;
Matches 1545; Conservative 0; Mismatches 10; Indels 10; Gaps 3;
QY 189 GAAAGATTATCACTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGGAGAGTTGC 248
DB 142 GAAAGATTATCACTTTGAATATACGGAATGTGATAGCAGTGGCTCCAGGTGGAGAGTTGC 201
QY 249 CATTCCAAATTCGAGTGGACTGCTGGCCTGCTGACCCAGTCAGAGGCAAGAATG 308
DB 202 CATTCCAAATTCGAGTGGACTGCTGGCCTGCTGACCCAGTCAGAGGCAAGAATG 261
QY 309 CACTTCTCCTGCTCTGGAGATATCTAGAAATGAAGAACCCAGGTATGAGTAAAGT 368
DB 262 CACTTCTCCTGCTCTGGAGATATCTAGAAATGAAGAACCCAGGTATGAGTAAAGT 321
QY 369 TGGTGAAGCAGCTATCTCTGGAGTGGCATCAAAATTTGATGATGGATGAATGGC 428
DB 322 TGGTGAAGCAGCTATCTCTGGAGTGGCATCAAAATTTGATGATGGATGAATGGC 381
QY 429 GCGAGATTTTCTACATCGCAACATTCATGGACACTGTGGTGGGCCCTTCTGACAGCAG 488
DB 382 GCGAGATTTTCTACATCGCAACATTCATGGACACTGTGGTGGGCCCTTCTGACAGCAG 441
QY 489 GCGAGAGGCTGTAAACACTCTTCTGGATCCCTCGTGGAACTACATAGATCTAATCG 548
DB 442 GCGAGAGGCTGTAAACACTCTTCTGGATCCCTCGTGGAACTACATAGATCTAATCG 501
QY 549 TGATGACTGACGGTGTCTTTGATCTATGCTGTGCACCTTAAGAGTCAAGGCTATGCTTT 608
DB 502 TGATGACTGACGGTGTCTTTGATCTATGCTGTGCACCTTAAGAGTCAAGGCTATGCTTT 561
QY 609 CTTTGAGTACCACTATGTGCAACAACATCTCTTTGAGTCTTTTATTCAAAATGATCA 668
DB 562 CTTTGAGTACCACTATGTGCAACAACATCTCTTTGAGTCTTTTATTCAAAATGATCA 621
QY 669 GTGCCAGGAGATGGACACCACTGACAAAGTGGTAAACCTTACAGCAATGGAGAATG 728
DB 622 GTGCCAGGAGATGGACACCACTGACAAAGTGGTAAACCTTACAGCAATGGAGAATG 681
QY 729 GGGCTCTCAATCTGTAATGCTGAAATCAGGCACAACTACTCTACTGGAGAACTACAGG 788
DB 682 GGGCTCTCAATCTGTAATGCTGAAATCAGGCACAACTACTCTACTGGAGAACTACAGG 741
QY 789 CATCTCTTATGGGTCTTAAGCGGCTCAAGCCTGTGCTGGTAAATAATATCACAATTTGAAG 848
DB 742 CATCTCTTATGGGTCTTAAGCGGCTCAAGCCTGTGCTGGTAAATAATATCACAATTTGAAG 801
QY 849 GGTGGGCTACACATCAAGATGTTTCTTGCAGCCAGGCACATTCAGCAACAACCCAGG 908
DB 802 GGTGGGCTACACATCAAGATGTTTCTTGCAGCCAGGCACATTCAGCAACAACCCAGG 861

QY 909 TTCATTCAACTGCCAGGTGTGTCCAGAAACACCTATTCTGAGAAAGGAGCCAAAGAATG 968
DB 862 TTCATTCAACTGCCAGGTGTGTCCAGAAACACCTATTCTGAGAAAGGAGCCAAAGAATG 921
QY 969 TATAAGGTGTAAGACGACTCTCAATTTTC-----AGGATCCAGTGTGATACAGAGCG 1022
DB 922 TATAAGGTGTAAGACGACTCTCAATTTTCAGAGAAAGATCCAGTGTGATACAGAGCG 981
QY 1023 CCTTCCCTGTACCACAAAGACTATTTCAGATCCATCTACTCCATGTGATGAAGAAGAAA 1082
DB 982 CCTTCCCTGTACCACAAAGACTATTTCAGATCCATCTACTCCATGTGATGAAGAAGAAA 1041
QY 1083 GACACAGATAATGTACAGTGGATAGAGCCCAAAATCTCCGGGAGGATCTCACAGATGC 1142
DB 1042 GACACAGATAATGTACAGTGGATAGAGCCCAAAATCTCCGGGAGGATCTCACAGATGC 1101
QY 1143 TATTAGATTGCCCTTCTGGAGAGAAGAGATTTGTCCGCCCTTGCACCCCTCGATTTTA 1202
DB 1102 TATTAGATTGCCCTTCTGGAGAGAAGAGATTTGTCCGCCCTTGCACCCCTCGATTTTA 1161
QY 1203 TAACAATGGATCATCTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1262
DB 1162 TAACAATGGATCATCTTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1221
QY 1263 AGAATGTAGACCATCTCCAGCAGGAGGAGGCTGCACCTTGGCTTTGAAATATAAATGGTG 1322
DB 1222 AGAATGTAGACCATCTCCAGCAGGAGGAGGCTGCACCTTGGCTTTGAAATATAAATGGTG 1281
QY 1323 GAATGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1382
DB 1282 GAATGTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1341
QY 1383 TGAATGAATGTTGGAGGTGGCTGGAGATCATATCCAGAGTGGGCTGGAGGTCTTGA 1442
DB 1342 TGAATGAATGTTGGAGGTGGCTGGAGATCATATCCAGAGTGGGCTGGAGGTCTTGA 1401
QY 1443 CAATGATACCTGATCTTAACTTGCATATCCAGATTTAAACCCACCAACATCTATGAC 1502
DB 1402 CAATGATACCTGATCTTAACTTGCATATCCAGATTTAAACCCACCAACATCTATGAC 1461
QY 1503 TGGAGCCAGGGTCTGAACTAGGAAGAATAACATTTGTCTTTGAGACCCCTCTGTTTACG 1562
DB 1462 TGGAGCCAGGGTCTGAACTAGGAAGAATAACATTTGTCTTTGAGACCCCTCTGTTTACG 1521
QY 1563 TGACTGTGTTTGTACTTTCATGATGATTAATAGAAAAGTACAAAATGTGTTAGAAATC 1622
DB 1522 TGACTGTGTTTGTACTTTCATGATGATTAATAGAAAAGTACAAAATGTGTTAGAAATC 1581
QY 1623 GTGGGTGGAACCAAGAAACAAAGCTTACACCCATATCATCTTCAAGAATGCCAATTT 1682
DB 1582 GTGGGTGGAACCAAGAAACAAAGCTTACACCCATATCATCTTCAAGAATGCCAATTT 1641
QY 1683 TACATTTACATGGG--CATTTCCAGAGAACTAAT--CAGGGTCAAGATTAATAGACGGTTCA 1738
DB 1642 TACATTTACATGGGCAATTTCCAGAGAACTAATTTAGGGTCCAAAGATAATAGACGGTTCC 1701
QY 1739 TCAAT 1743
DB 1702 NCCAT 1706

RESULT 4

AAC84886

ID AAC84886 standard; cDNA; 1508 BP.

XX

AC AAC84886;

XX

DT 20-APR-2001 (first entry)

XX

DE Human SEC5 nucleic acid sequence (clone ID 1795045.0.61).

XX

KW SECX; cytostatic; gynecological; gene therapy; screening assay; human; chromosomal mapping; forensic biology; cell proliferation; cancer;

KW cell differentiation; immune associated disorder; gestational disease;
 KW SEC5; ss.
 XX
 OS Homo sapiens.

XX
 FH Key Location/Qualifiers
 FT CDS 226..1461
 FT /*tag= a
 FT /product= "SEC5"

XX WO200078802-A2.

XX 28-DEC-2000.

XX 23-JUN-2000; 2000WO-US17328.

XX 23-JUN-1999; 99US-0140584.

XX 20-JUL-1999; 99US-014722.

XX 16-SEP-1999; 99US-0154520.

XX 22-JUN-2000; 2000US-0604286.

XX (CURA-) CURAGEN CORP.

XX PI Shinkets RA, Fernandes E, Vernet C, Yang M, Boldog FL;
 PI Herrmann JL;

XX DR WPI; 2001-071385/08.

XX DR P-PSDB; AAB48372.

XX Polynucleotides encoding SECX proteins useful for treating disease
 PT characterized by an aberrant level of cell proliferation and/or
 PT differentiation like cancer or immune associated disorders.

XX PS Claim 3; Fig 5; 132pp; English.

XX CC The invention relates to human SECX polypeptides and polynucleotides
 CC encoding them. The SECX polypeptides can be expressed by standard
 CC recombinant methodology. The SECX polypeptides are useful for treating
 CC or preventing a SECX-associated disorder. The invention is useful in
 CC screening assays; detection assays (e.g. chromosomal mapping, cell and
 CC tissue typing, forensic biology); predictive medicine (diagnostic assays,
 CC prognostic assays, monitoring clinical trials, and pharmacogenomics); and
 CC methods of treatment (e.g. therapeutic and prophylactic), especially
 CC disorders characterized by aberrant cell proliferation and/or
 CC differentiation like cancer or immune associated disorders or gestational
 CC disease. The present sequence represents a SEC5 nucleic acid sequence.

XX SQ Sequence 1508 BP; 444 A; 306 C; 355 G; 402 T; 1 other;

Query Match 34.9%; Score 1182.6; DB 22; Length 1508;
 Best Local Similarity 98.4%; Pred. No. 2.5e-300;
 Matches 1228; Conservative 0; Mismatches 10; Indels 10; Gaps 3;

QY 506 ACTCTCTTGGATCCCTCGTGAACACTACATAGATCTAATCGTGTACTGCACGGGT 565

DB 230 ACTCTCTTGGATCCCTCGTGAACACTACATAGATCTAATCGTGTACTGCACGGGT 289

QY 566 CTTTGTATCTATGCTGTGCACCTTAAGAAGTCAGGCTATGCTCTTTGGAGTACCACTATG 625

DB 290 CTTTGTATCTATGCTGTGCACCTTAAGAAGTCAGGCTATGCTCTTTGGAGTACCACTATG 349

QY 626 TCGACAAACATCTCTTTTGAGTTCTTTATTTCAAAATGATCAGTGCCACGAGATGGACA 685

DB 350 TCGACAAACATCTCTTTTGAGTTCTTTATTTCAAAATGATCAGTGCCACGAGATGGACA 409

QY 686 CCACCACTGACAGTGGGTAACACTTACAGACAATGGAGAATGGGGCTCTCATTTCTGTAA 745

DB 410 CCACCACTGACAGTGGGTAACACTTACAGACAATGGAGAATGGGGCTCTCATTTCTGTAA 469

QY 746 TCGTGAATCAGGCACAAACATCTCTACTGAGACTACAGCATCTTATGGGTTCTA 805

DB 470 TCGTGAATCAGGCACAAACATCTCTACTGAGACTACAGCATCTTATGGGTTCTA 529

QY 806 AGCGGTCAAGCCTGTCTGTGTAAAAATATCACAAATTCAGGGGTGCGTACACATCAG 865
 DB 530 AGCGGTCAAGCCTGTCTGTGTAAAAATATCACAAATTCAGGGGTGCGTACACATCAG 589
 QY 866 AATGTTTCTTTCGCAAGCCAGGCACATTCAGCAAAACACAGGTTTCATTCACCTCCAGG 925
 DB 590 AATGTTTCTTTCGCAAGCCAGGCACATTCAGCAAAACACAGGTTTCATTCACCTCCAGG 649
 QY 926 TGTGTCCAGAAACACCTATTCTGAGAAAGGAGGACCAAGAAATGTAAGGTTGTAAGACG 985
 DB 650 TGTGTCCAGAAACACCTATTCTGAGAAAGGAGGACCAAGAAATGTAAGGTTGTAAGACG 709
 QY 986 ACTCTCAATTTTC-----AGGATCCAGTGTGTACAGAGCGCCCTCCCTGTACCACAA 1039
 DB 710 ACTCTCAATTTTCAGAGGAAGGATCCAGTGTGTACAGAGCGCCCTCCCTGTACCACAA 769
 QY 1040 AAGACTATTTCAGATCCATCTCTGATGTGAGAGGAAAGACACAGATAATGTACA 1099
 DB 770 AAGACTATTTCAGATCCATCTCTGATGTGAGAGGAAAGACACAGATAATGTACA 829
 QY 1100 AGTGATAGAGCCCAAAATCTCGCGGAGGATCTCACAGATGCTATTAGATTGCCCTTT 1159
 DB 830 AGTGATAGAGCCCAAAATCTCGCGGAGGATCTCACAGATGCTATTAGATTGCCCTTT 889
 QY 1160 CTGGAGAGAAGGATTTCTCGCCTTGCAACCCCTGGATTTTATAACATGATCATCTT 1219
 DB 890 CTGGAGAGAAGGATTTCTCGCCTTGCAACCCCTGGATTTTATAACATGATCATCTT 949
 QY 1220 CTTGGCATCCCTGCTCTCTGGAACATTTTCAGATGGAACCAAGAAATGTAGACCATGTC 1279
 DB 950 CTTGGCATCCCTGCTCTCTGGAACATTTTCAGATGGAACCAAGAAATGTAGACCATGTC 1009
 QY 1280 CAGCAGGAACGAGCCTGCACCTTGGCTTTGAATATAAATGSGGAATGCTCTTCCTGGCA 1339
 DB 1010 CAGCAGGAACGAGCCTGCACCTTGGCTTTGAATATAAATGSGGAATGCTCTTCCTGGCA 1069
 QY 1340 ACATGAAAACCTTCCTGCTTCAATGTTGGGAATTCAGAGTCGATGGAATGAATGTTGGG 1399
 DB 1070 ACATGAAAACCTTCCTGCTTCAATGTTGGGAATTCAGAGTCGATGGAATGAATGTTGGG 1129
 QY 1400 AGTGGCTGGAGATCATATCCAGATGGGGCTGGAGGTTCTGACAAATGATACCTGATCT 1459
 DB 1130 AGTGGCTGGAGATCATATCCAGATGGGGCTGGAGGTTCTGACAAATGATACCTGATCT 1189
 QY 1460 TAACTTGATATCCAGGATTTAAACCAACCAACATCTATGACTGGAGCCACGGGTTCTG 1519
 DB 1190 TAACTTGATATCCAGGATTTAAACCAACCAACATCTATGACTGGAGCCACGGGTTCTG 1249
 QY 1520 AACTAGGAAGATAACATTTGTCTTTTGAGACCCCTCTGTTACGCTGACTGTGTTTGTACT 1579
 DB 1250 AACTAGGAAGATAACATTTGTCTTTTGAGACCCCTCTGTTACGCTGACTGTGTTTGTACT 1309
 QY 1580 TCATGGTGGATATTATAGAAAAAGTACAAATGTGGTAGAATCGTGGGTGGAACCAAG 1639
 DB 1310 TCATGGTGGATATTATAGAAAAAGTACAAATGTGGTAGAATCGTGGGTGGAACCAAG 1369
 QY 1640 AAAAAACACTTACACCCCATATCATCTTCAAGAATGCAACTTTTACATTTACATTTACAT 1697
 DB 1370 AAAAAACACTTACACCCCATATCATCTTCAAGAATGCAACTTTTACATTTACATTTACAT 1429
 QY 1698 ATTCCAGAGAACTAAT--CAGGGTCAAGATAATAGACGGTTTCATCAAT 1743
 DB 1430 TTCCCAGAGAACTAATTCAGGGTCCCAAGATAATAGACGGTTCCNCCAT 1477

RESULT 5
 AAA95442

ID AAA95442 standard; cDNA; 3280 BP.

XX AAA95442;

XX AC AAA95442;

XX 12-FEB-2001 (first entry)

XX

Db 1563 CAACACTCTCTGGAGACGTGGAAGGTTCCAAAGGCAACACAGTCTATACCTATCATCAT 1622
Qy 1665 CTTCAAGATCAACACTTTTACATTTTACATGGGCATTTCAGAGCAACTAATCAGGTCAGA 1724
Db 1623 TCAGAGAACACTACACGAGCTTCACCTGGGCCCTTCAGAGGACCACTTTTCATGAGGC 1682
Qy 1725 TAATAGACGGTTCAATCAATGACATGGTGAAGATTATTTCTATCAGAGCCACTAATGCGAGT 1784
Db 1683 AAGCAGGAAGTACACCAATGAGCTTGCCAAAGATCTACTCCATCAATGTCCACCAATGTAT 1742
Qy 1785 TCATGGGTGGGCTCTCATGCGGTGCGCTGTCGCCCTGGTTCTGAACAGTCGGGTTCATC 1844
Db 1743 GAATGGCTGGGCTCTCTACTGCGCTCCCTGTGCCCTAGAACCTCTGTATGTGGGCTCCTC 1802
Qy 1845 GTGTGTCCCTGCCCTCCAGGCCACTACATTCAGAAAGAAACCAACCACTGTCAGGAATG 1904
Db 1803 CTGCACCTTGTCTCTGCTGTTACTATATTGACCGAGATTCCAGAACCTGCCACTCCTG 1862
Qy 1905 TCCACCTGACACTACCTGTCCATACATCAGGTCTATGGCAAGAGGCTTGTATTCCCATG 1964
Db 1863 CCCCCCTAACCAATTCGAAAGCCACACAGCTTATGGTCCAGGCTGTGTGCCCTG 1922
Qy 1965 CGGGCTGGGAGTAAACAAATCAGACCAATTCGGTTTGGCTATAGTGACTGCTTTTCTA 2024
Db 1923 TGGTCCAGGACCAAGAACCAAGATCCACTCTCTGTGCTACAATGATTGCACCTTCTC 1982
Qy 2025 CCATGAAAGAAATCAGATTTTGCATCTATGACTTACCACTCAGCAGTGTGGGCTC 2084
Db 1983 AGCAACACTCCACAGAGACTTTCACTCAACATTCCTCGGCTTTGGCAACACCGTCA 2042
Qy 2085 ATTAATGAATGGCCCGCCAGCTTCACCTCCAAAGAAACAAATACTTCCATTTCTCAATAT 2144
Db 2043 TCTTCTGGAGGCCAAGCTTCACITCCAAAGGTTGAAATACTTCCATCCTTTACCT 2102
Qy 2145 CAGTTTATGGGATGAGGGAAGAAGATGCTCTGTGTACCAACAATATAACAGACTT 2204
Db 2103 CAGTCTCTGTGGAAACACAGGTPAGGAAATGTCTGTGTACCGCAATGTCACTGACCT 2162
Qy 2205 TACAGTAAAGAAATAGTGGCAGGTCAGATGATTACACAAATTTGGTAGGGCAATTTGT 2264
Db 2163 CCGGATTCCTGAGGTGAGTCAGG-----TTCCTCAATCTATCACAGCCTACGT 2213
Qy 2265 ATGCCAGTCAACAATTAATTCCTCTGAAAGTAAAGGTTTCCGAGCAGCCTTATCATCACA 2324
Db 2214 CTGCCAGGAGTCATATCCCCCAGAGGTGACAGCTACAGGCGGGGTTTCTCTACA 2273
Qy 2325 ATCCATCATCTGGCAGATACATTCATAGGAGTCACAGTTGAAACACATTTGAAATAAT 2384
Db 2274 GCCTGTACGCTTGTGTGACGACTTATTGGGTGACACAGATATGACTCTGGATGGAT 2333
Qy 2385 TAATATAAAGAAATATGTTCCAGTTTCCAAACAGCCAAATACCAGATGTCAATTTCTT 2444
Db 2334 CACCTCCCGAGCTGACATTTCCACCTCGAGTCTTGGGAATACCGAGGTGATCTTCT 2393
Qy 2445 TTATAAGTCTTCTACAGCAACAATCTTGTATTATGGCCGATCAACTGCTGTGAAAT 2504
Db 2394 TTATAGTCCAATGATGTGACCCAGTCTCTGAGTTCTGGGAGATCAACCACTCCGGT 2453
Qy 2505 GAGGTGTAATCTACTAATATGAGGAGGAGTGAATTCAGTCCCGCAGCAAGTGGCCAGC 2564
Db 2454 CAGGTGCACTCACAGAAACTGTCCCTGGAAAGTTGTCTGTCCAGGAACCTGTCAGA 2513
Qy 2565 AGGTACCTGTGATGGGTGACCTTCTATTCTCTGGGAGAGTGTGAGCTTGGCCCTCT 2624
Db 2514 TGGGACCTGTGATGGCTGCACTTCCACTTCTGTGGGAGAGCGGGTGTCTTGGCCGCT 2573
Qy 2625 GTGTACGGAGCATGACTTCCATGAGATTGAGGAGCCCTGCAAGAGAGGATTTTCAGGAAC 2684
Db 2574 CTGCTCAGTGGCTGACTACCATGCTATCGTCAGCAGCTGTGTGGCTGGATCCCAAGAC 2633
Qy 2685 CTTGTATGTGGAATGAACCTTAATGGTCATTAAGGAATTTCTTTGGCTGAGAAAA 2744

Db 2634 TACTTACGTGTGGCGAGAACCAAGCTATGCTCTGTGGCATTTCTCTGCTGAGCAGAG 2693
Qy 2745 GTTGGCAACCTCTGAAACGTTGACTTTTGGCTGAGGTGGGAGCGGTGTGGGACTTT 2804
Db 2694 AUTCACACTCTGCAAAACCATAGATTTCGGCTGAAAGTGGGCATCTCTGCAGGCACCTG 2753
Qy 2805 TACTGCGGTTTGTGGTGGCTGTGACCTGTGACTTCTGGAAAAAATCAAAAAA 2864
Db 2754 TACTGCCATCTCTGCTCACCGTCTTGACCTGTGACTTTTGGAAAAAATCAAAAAA 2813
Qy 2865 GAAGACCATTTGAATCTCTTCA 2887
Db 2814 GTACAAGTACTCCAAGCTGGTGA 2836

RESULT 6
AAF89765
ID AAF89765 standard; cdNA; 3331 BP.
AC AAF89765;
XX 23-JUL-2001 (first entry)
DT 23-JUL-2001 (first entry)
XX Nucleotide sequence of a human protein expressed in tumour cells.
DE Tumour cell; immunological disease; autoimmune disease; cancer;
KW infection; ss.
XX Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 45..3086
FT /*tag= a
FT /product= "tumour expressed protein"
FT sig_peptide 45..167
FT /*tag= b
XX WO200131003-A1.
XX 03-MAY-2001.
XX 30-OCT-2000; 2000WO-FR03032.
XX 29-OCT-1999; 99FR-0013629.
XX (FABR) FABRE MEDICAMENT SA PIERRE.
XX Delneste Y, Magistrelli G, Jeannin P, Bonnefoy J;
XX WPI: 2001-328651/34.
XX P-PSDB; AAB83845.
XX New nucleic acid, expressed in tumours and lymphoid tissue is useful for
XX identifying agents for treating tumours and autoimmune disease .
XX Claim 1; Page 43-48; 85pp; French.
XX
XX The present sequence encodes a human protein expressed in tumour cells.
XX The polynucleotide is useful for screening cDNA/genomic DNA banks and
XX for cloning isolated DNA; identifying mutant forms of the gene that
XX encodes a human protein, where the mutations are associated with
XX abnormal gene expression, or promoters and regulators of the gene,
XX particularly for diagnosis; for recombinant expression of the derived
XX protein; as probes and primers for detection and amplification; and
XX as antisense therapeutics. The tumour expressed protein is useful for
XX raising specific antibodies and to screen agents that modulate its
XX activity, bind to it or interact with it. These agents are potentially
XX useful for treatment or prevention of diseases associated with abnormal
XX expression/activity of the protein, particularly immunological diseases
XX (autoimmune diseases and cancer) or viral, bacterial, fungal or parasitic
XX infections.
XX Sequence 3331 BP; 822 A; 945 C; 813 G; 751 T; 0 other;

Query Match 26.6%; Score 903; DB 22; Length 3331; Best Local Similarity 60.1%; Pred. No. 1.4e-226; Matches 1636; Conservative 0; Mismatches 1060; Indels 27; Gaps 7;									
Qy	171	GCTTCCTCTTCCAGGAGAAAGATATACATTTTGAATATAGGAAATGTATAGACAGTGG	230						
Db	179	GCTTCATGCTCGCAAGAGCTCTGAGTACCACTATGATGATACAGGGCTGTGACAGCAGCG	238						
Qy	231	CTCCAGGTGGAGATTGCCATTCCAAATTTCTGCAGTGGACTGCTCTGGCCCTGCCGTGACCC	290						
Db	239	TTCCAGGTGGAGGTCCCGCTGCCGATACCCGGCCGTGTGCACAGCCTGCCGTGACCC	298						
Qy	291	AGTGAGAGCAAGAATCTCTCTCTGCTTCTGAGAGTATCTAGAAATGAAGAA	350						
Db	299	CGTCAAGGCGCAGGAGTCTCTCTCTCTGCAAGCGCGGAGTTTCTCGATATGAAGA	358						
Qy	351	CCAGGTATCGAGTAAAGTGGTGAAGGACCTATTTCTTTGGGAGTGGGATCAAAATTTGA	410						
Db	359	CCAGTCAATGTAAAGCCATGCGCTGAGGGCGCTACTCCCTCGGCACAGGATTCGGTTTGA	418						
Qy	411	TGAATGGGATGAATGCCCGCAGGATTTTCTAAATCGCAACATTCATGGACACTGTGT	470						
Db	419	TGAGTGGGATGAGTGGCCCATGGCTTCCAGCCTCTCAGCCAACATGGAGCTGGATGA	478						
Qy	471	GGGCCCTTCTGACAGCAGGCCAGCGCTGTAAACAACTCTTCTTGGATCCCTCGTGGA	530						
Db	479	CAGTGTCTGAGTCCA---CGGGGAACGTACTTCTCCAAAGTGGTTCCCGGGCGCA	535						
Qy	531	CTACATAGAATCTAATCGTGATGACTGACGGTGCTTTTGATCTATGCTGTGCACCTTAA	590						
Db	536	CTACATCGGCTCCACAGGAGGAATGCAAGCCACACTGATGAGCGCTCAACCTGAA	595						
Qy	591	GAAGTCAGGCTATGCTTTTGTAGTACCAGTATGTCGACAAACATCTCTTTGAGTT	650						
Db	596	GCAATCTGCCACCGTTAACTTGAATACTATCCAGACTCCAGCATCATCTTTGAGTT	655						
Qy	651	CTTTATTCAAATGATCATGTCGAGAGATGGACACCACTGACAAATGGGTAAACT	710						
Db	656	TTTCGTTCAGATGACAGTGGCGGAGCCCAATGCAGATGACTC---CAGGTGGATGAAGAC	712						
Qy	711	TACAGACAATGGAATGGGGTCTCATTTCTGTATGCTGAAATCAGGCACAAACATACT	770						
Db	713	CACAGAGAA---AGATGGGAATTCACAGTGTGGAGTAAATTCGAGGCAATATGCTCT	769						
Qy	771	CTACTGGAGAACTACAGGATCCTTATGGGTTCTAAGCGGTCAAGCCTGTGCTGTAA	830						
Db	770	CTATTGGAGAACACAGCCTTCTCAGTATGAGCAAGTACCAGCCTGTGCTGTGAG	829						
Qy	831	AAATATCAAAATGAAGGGTGGCGTACACATCAGAACTATTTTCTTTGCAAGCCAGGCAC	890						
Db	830	AAACATTTGTCATAAACAGGGTGGCGCTACACTTTCAGAAATGCTTCCCTGCAAACTGGCAC	889						
Qy	891	ATTACAGCAACAACAGGTTTCACTCAACTGCCAGTGTGTCGCCAGAAACACCTATTCTGA	950						
Db	890	GTATGCAGACAGCAGGGTCTCTTTCTGCAAACTTTGCCAGGCAACTTATTTCAA	949						
Qy	951	GAAAGAGCAAGAAGATGTATAGGTGTAAA---GACGACTCTCAATTTTCAGGATCCAG	1007						
Db	950	TAAAGGAGAACTTCTTGCCACAGTGTGACCTTGACAAATCTCAGAAAGGATCTTC	1009						
Qy	1008	TGAGTGTACAGAGCCCTCCCTGTACCAACAAGAACTATTTTCCAGATCCATCTCCATG	1067						
Db	1010	TTCTCTGATACGTGGCGGAGCTTGCACAGACAAAGATTTATTTCTACACACACAGCGCTG	1069						
Qy	1068	TGATGAAGAGAAACACAGATATGTATCAAGTGGATAGAGCCCAAAATCTGCCGGA	1127						
Db	1070	CGATGCCAACGGAGACACAACTCATGTACAAATGGGCCCAAGCCGAAATCTGTAGCA	1129						
Qy	1128	GGATCTCACAGATGCTATTAGATTGCCCTCTTCTGGAGAGAAAGGATTTGTCGCGCTTG	1187						
Db	1130	GGACCTTGAGGGGGAGTCAAGCTGCCCTGCTGTGTGTGAGAGCCCACTGCCACCCCTG	1189						

Qy	1188	CAACCCCTGGATTTTATPACAAATGGATCATCTTTCTTGCCATCCCTCTCTCTGGAACATTT	1247						
Db	1190	CAACCCAGGCTTCTTCAAAACCAACAACAGCAGCTGCCAGCCTGCCCATATGTTGTCCTTA	1249						
Qy	1248	TTTCAGATGAACCAAGAATGTAGACCATTGTCCAGCAGGAACGGAGCCTGACATTTGGCTTT	1307						
Db	1250	CTTCCAATGG---CTCAGACTGTACCGCTGCCCTGCGAGGACTGAACCTGCTGTGGGATTT	1306						
Qy	1308	TGAATATAATGTTGGAATGTCTTCTTCCGCAACATTTCTCTCAATTTCTTAAATTTGG	1367						
Db	1307	TGAATATAAATGTTGGAACACGCTGCCCAACAACATGGAACACGCGTTCTCAGTGGGAT	1366						
Qy	1368	GAATTCAAAGTCGATGGAATGAATGGTTGGAGGTGGCTGGAGATCATATCCAGAGTGG	1427						
Db	1367	CAACTTCGAGTACAAGGCGTACAGGCTGGAGGTGGCTGGTGTGATCATATTTACACAGC	1426						
Qy	1428	GGCTGGAGGTTCTGACAATGATTACCTGATCTTAACTTTGCATATCCAGGATTTAAACC	1487						
Db	1427	TGCTGGAGCCTCAGACAATGACTTTCATGATTTCTACTCTGGTTGTGCCAGGATTTAGACC	1486						
Qy	1488	ACCAACATCTATGACTGGAGCCAC---GGGTTCTGAACTAGGAAGAATAAACATTTGCTTT	1544						
Db	1487	TCCGCACTCGGTGATGCGACACACAGAGATTAAGAGGTGGCCAGANTCNCATTTGCTTT	1546						
Qy	1545	TGAGACCCCTCTGTTCAGCTGACTGTGTTTGTACTTTTCATGTTGGATTAATAAGAAAAAG	1604						
Db	1547	TGAGACCCCTCTGTCTGTGAACCTGTGAGCTTACTTTTCATGGTGGTGTGAATTTCTAGGAC	1606						
Qy	1605	TACAAATGTGGTAGAATCGTGGGTGGAACCAAGAAACAAAGCTTACACCAATATCAT	1664						
Db	1607	CAACACTCTCTGGAGACGCTGGAAAGGTTCCAAAGGCAACACAGTCTTATACCTTACATCAT	1666						
Qy	1665	CTTCAAGAATGCAACTTTTACATTTTACATGGGCATTTCCAGAGAACTTAATCAGGGTCAAGA	1724						
Db	1667	TGAGGAGAACACTACCACGAGCTTACCCTGGGCTTCCAGAGGACCACTTTTCATGAGGC	1726						
Qy	1725	TAAATAGACGGTTTCAATCAATGATGATGAAGATTTATTTCTATCAGAGCAGCTAATGCGAT	1784						
Db	1727	AAGCAGGAAGTACACCAATGAGCTTGCAGAGATCTACTCCATCAATGTCAACCAATTTAT	1786						
Qy	1785	TGATGGGTGGGTGCTCATGCGTGCCTGTGCCCTCTCTGACAGCTGCGGTTTCATC	1844						
Db	1787	GAATGGGTGGGCTTCTACTGCGCTCCCTGTGCCCTTAGAAGCTCTGATGTGGGCTCTC	1846						
Qy	1845	GTGTGTCCCTTGCCTTCCAGGCGCACTTACATTTGAGAAAGAAACCAACAGTGAAGGAATG	1904						
Db	1847	CTGCCACTCTTGTGCTGCTGTTACTATATATGACCGAGATTTCAGGAACCTGCCACTCTC	1906						
Qy	1905	TCCACTTGACACTTACTGTCCATACATCAGGTCTATGGCAAGAGGCTTGTATTCATG	1964						
Db	1907	CCCCCTTACACAAATTTCTGAAAGCCCAACAGCCTTATGGTGTCCAGGCTGTGTGCCCTG	1966						
Qy	1965	CGGGCTGGGAGTAAACAAATCAGGACCAATTCGGTTTCTATAGTACTGCTTTTCTTA	2024						
Db	1967	TGTTCCAGGAGCAAGAACACAGATCCACTCTCTGTGCTACAATGATTTGCACTTTCTC	2026						
Qy	2025	CGATGAAAAGAAAATTCAGATTTTGCACTTATGACTTTAGCAACCTCAGAGTGTGGGCTC	2084						
Db	2027	ACGCAACACTCCAAACCCAGGACTTTCAACTTACAACTTCTCCGCTTTGCAACACCCGTCAC	2086						
Qy	2085	ATTAATGAATGGCCCCCAGCTTCACTTCCAAAGGAACAAATACTTCCATTTCTTCAATAT	2144						
Db	2087	TCTTGTGGAGGGCCAAAGCTTCACTTCCAAAGGGTTTGAATTTCCATCTCACTTTACCT	2146						
Qy	2145	CAGTTTATGTGGCATGAGGGAAGAGATGGCTCTCTGTGACCAACAAATATACACACTT	2204						
Db	2147	CAGTCTCTGTGAAACCCAGGATAGGAAATGTCTGTGTGCAACGACAATGTCACTGACCT	2206						
Qy	2205	TACAGTAAAGAAATAGTGGCAGGCTCAGATGATTAACAAAATTTTGGTAGGGCATTTGT	2264						
Db	2207	CGGGATTTCTGTGAGGGTGTGTCAGG-----TTTCTCCAATCTATCACAGCCCTACGT	2267						
Qy	2265	ATGCCAGTCAACAAATTTATTTCTTCTGAAAGTTAAGGGTTTCCGAGGAGCCCTTATCATCACA	2324						

Db 2258 CTGCCAGGCGATCATATCCCCAGAGGTGACAGGCTACAGGCGGGTTTCTCTACA 2317
Qy 2325 ATCCATATCTTCGGAGATACATTCATAGGAGTACAGTGTGAACACATTTGAAATAT 2384
Db 2318 GCCTGTACGCTTGTGATCGACTTATTTGGGGTGACACAGATATGACTCTGGATGGAAT 2377
Qy 2385 TAATATAAAGAGATATGTTCCCGATTTCCAAAGCCCAATACAGATGTCATTTCTT 2444
Db 2378 CACCTCCACGCTGACTTTTCCACCTGGAGTCTTGGGAATACCGGAGTGATCTCTT 2437
Qy 2445 TTATAAGCTTCTTACAGCAACATCTTGTATTATTAATGGCCGATCAACTGCTGTGAAAT 2504
Db 2438 TTATAGTTCCTAATGATGACCCAGTCTGTCAGTCTTGGGAGATCAACACCATCCGGGT 2497
Qy 2505 GAGGTGTAATCTTACTAATCTGAGCAGGAGTATTTCACTGCTCCACAGCAAGTCCGAC 2564
Db 2498 CAGGTGCACTCCACAGAAACTGTCCCTGGAAAGTTTGTGCTGCCAGGAAGTGTCTCAGA 2557
Qy 2565 AGGTACCTGTGATGGGTGAGTCTTATTTCTGTGGGAGAGTGTGAAGCTTGGCCCTCT 2624
Db 2558 TGGACCTGTGATGGCTGCACTTCCACTTCTGTGGGAGAGCGGCTGCTTGGCCGCT 2617
Qy 2625 GTGACGAGCATGACTTCCATGATGAGATTGAGGAGCCCTGCAAGAGAGGATTTTCAGGAAC 2684
Db 2618 CTGCTCAGTGTGACTACCTGCTATCTGTCAGCAGCTGTGTGGCTGGGATCCAGAGAC 2677
Qy 2685 CTTGTATGTGTGGATGAACCTAATAGTGTGATTAAGGAATTTCTTTGCTGTGAGAAAA 2744
Db 2678 TACTTACGTGTGGGAGAGAACCAAGTATGCTCTGGTGGCAATTTCTCTGCTGTGACGAG 2737
Qy 2745 GTTGGCAACCTGTGAACCGTTGACTTTTGGCTGAAGGTGGGAGCCGCTGTGGGAGCTTT 2804
Db 2738 AGTACCATCTGCAAAACCATAGATTCTTGCTGAAAGTGGGATCTCTGCAAGCACCTG 2797
Qy 2805 TACTGCCCTTTTGTGCTGTGCTGACTTCTGACCTGCTTCTTGGAAAAAGATCAAAAGAAAA 2864
Db 2798 TACTGCCATCTGCTCAACGCTTGTGACCTGCTACTTTTGGAAAAAGATCAAAACTAGA 2857
Qy 2865 GAAGACCATTTTGAATCTGTTCA 2887
Db 2858 GTACAGTACTCCAGCTGGTGA 2880

RESULT 7
AAF28030
ID AAF28030 standard; DNA; 3334 BP.
XX
AC AAF28030;
XX
DT 08-MAY-2001 (first entry)
XX
DE Human TR13 receptor coding sequence SEQ ID NO: 39.
XX
XX Human; tumour necrosis factor receptor; TR13; TR14; infection;
KW cancer; autoimmune disease; allergy; inflammatory disease;
KW graft rejection; apoptosis; cardiovascular disease; aneurysm; ds.
XX
OS Homo sapiens.
XX
PN WO200105834-A1.
XX
PD 25-JAN-2001.
XX
PF 14-JUL-2000; 2000WO-US19343.
XX
PR 16-JUL-1999; 99US-0144087.
PR 18-AUG-1999; 99US-0149450.
PR 20-AUG-1999; 99US-0149712.
PR 10-SEP-1999; 99US-0153089.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX

PI Ruben SM, Ni J, Young PE;
XX WPI: 2001-112682/12.
XX
PT Nucleic acids encoding 2 human tumor necrosis factor receptor
PT polypeptides (TR13) and (TR14), useful for the prevention, diagnosis,
PT and treatment of, e.g. cancers, acquired immune deficiency syndrome and
PT hypohidrotic ectodermal dysplasia -
XX
PS Claim 4; Page 394-398; 418pp; English.
XX
CC The present invention provides the protein and coding sequences of the
CC human tumour necrosis factor receptors TR13 and TR14. These sequences are
CC useful in the diagnosis and treatment of many diseases, including cancer,
CC autoimmune diseases, cardiovascular disorders, allergies,
CC neurodegenerative diseases, graft rejection, inflammation, aneurysms and
CC infections.
XX
SQ Sequence 3334 BP; 820 A; 952 C; 811 G; 751 T; 0 other;

Query Match 26.68; Score 901.4; DB 22; Length 3334;
Best Local Similarity 60.08; Pred. No. 3.7e-226;
Matches 1635; Conservative 0; Mismatches 1061; Indels 27; Gaps 7;

Qy 171 GCTTCTCTCTTGGCAGGAGAAAGATTATCACTTTGAATATACGGAATCTGATAGCAGTGG 230
Db 192 GCTTCACGCTTGCAAGAGTCTGAGTACCACCTATGATGACAGGCGTGTGACAGCACGGG 251
Qy 231 CTCAGGTGGAGATGTCATTTCCAAATTTCTGCAGTGGACTCTTGGCCCTGCCCTGACCC 290
Db 252 TTCCAGGTGGAGGTGCGCGTACCATACACCCGGCGCTGTGCCACGCTTGGCTGACCC 311
Qy 291 ACTGAGAGGCAAAAGTCCACTTCTCTCTGCTGCTCTGGAGAGTATCTAGAAATGAAGAA 350
Db 312 CGTCAAGGGCCACGAGTGTCTTCTCTCAACGCCGGGAGTTTCTGGATATGAAGGA 371
Qy 351 CCAGTATGTCAGTAAAGTGTGTGAAGCAGCCTATTCCTTGGGAGTGGCAGTGCATCAAAATTTGA 410
Db 372 CCAGTCACTGAAGCATCGCTGAGGCGCGCTACTCCCTCGCAGCAGCATTCGGTTTGA 431
Qy 411 TGAATGGGATGAATGCGCGGAGGATTTTCTAACATCGCAACATTCATGGACACTTGGT 470
Db 432 TGAGTGGGATGAGTGTCCCATGGCTTTGCCAGGCTCTCAGCCAAACATGGAGTGGATGA 491
Qy 471 GGGCCCTTCTGACAGCAGCAGCGCTGTAAACACTTCTTGGATCCCTCGTGGAAA 530
Db 492 CAGTGTCTGTGAGTCCA---CCGGGAAGTGTACTTCGCTCCAGTGGGTTCGCCGGGGCGA 548
Qy 531 CTACATAGAATCTAATCGTGTGACTGCACGCTGTCTTTGATCTATGCTGTGCACCTTAA 590
Db 549 CTACATCGCTTCAACACGGAGCATGACAGCCACACTGATGTACGCCGTCAACCTGNA 608
Qy 591 GAAGTCAGGCTATGCTTCTTTGAGTACAGTATGTCGACAAACAACTCTCTTTGAGTT 650
Db 609 GCAATCTGGCACCGTTAACTTCGAATACTACTATCCAGACTCCAGCATCATCTTTGAGTT 668
Qy 651 CTTATTCAAAATGATCAGTCCAGGAGATGGACACCACACTGACAAAGTGGTAAACT 710
Db 669 TTTGCTTCAGAAATGACCAATGCCAGTCCCAATGCAGATGACTC---CAGGTGGATGAAGAC 725
Qy 711 TACAGACAATGGAGAAATGGGCTCTCATTTCTGTAATCTGTAATCAGGCACAAACATACT 770
Db 726 CACAGAGAA---AGATGGGAATTCACAGTGTGGAGCTAAATCGAGCAATATGTCTT 782
Qy 771 CTACTGGAGAACTACAGCATCTCTTATGGGTCTTAAGCGGTCAAAGCCTGTGCTGGTAAA 830
Db 783 CTATTGGAGAACACAGCCTTCTCAGTATGGACCAAAAGTACCAGACCTGTGCTGTGAG 842
Qy 831 AAATATCAAAATGAGGGGTGGCGTACACATCAAGATGTTTCTTTCGACAGCCAGGCAC 890
Db 843 AAACATTCGCAATACAGGGGTGGCGTACACTTCAAGATGCTTCCCTCCAAACCTGGCAC 902
Qy 891 ATTACGAACAACACCAAGGTTTCAATTCACTGCGAGGTGTCTCCAGACAAACACCTATTCTGA 950

Human secretory and transmembrane; PRO; mammalian; cancer; lung; breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha; cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle; adipocyte; A-peptide; factor VIIa; gene therapy; ss.

Homo sapiens.

WO200140466-A2.

07-JUN-2001.

01-DEC-2000; 2000WO-US32678.

01-DEC-1999; 99WO-US28301.

01-DEC-1999; 99WO-US28634.

02-DEC-1999; 99WO-US28551.

02-DEC-1999; 99WO-US28564.

02-DEC-1999; 99WO-US28565.

09-DEC-1999; 99US-0170262.

16-DEC-1999; 99WO-US30095.

20-DEC-1999; 99WO-US30911.

20-DEC-1999; 99WO-US30999.

30-DEC-1999; 99WO-US31243.

06-JAN-2000; 2000WO-US00277.

06-JAN-2000; 2000WO-US00376.

11-FEB-2000; 2000WO-US03565.

18-FEB-2000; 2000WO-US04341.

18-FEB-2000; 2000WO-US04342.

22-FEB-2000; 2000WO-US04414.

24-FEB-2000; 2000WO-US04914.

24-FEB-2000; 2000WO-US05004.

01-MAR-2000; 2000WO-US05601.

20-MAR-2000; 2000WO-US07377.

21-MAR-2000; 2000WO-US07532.

30-MAR-2000; 2000WO-US08439.

17-MAY-2000; 2000WO-US13705.

22-MAY-2000; 2000WO-US14042.

30-MAY-2000; 2000WO-US14941.

02-JUN-2000; 2000WO-US15264.

10-NOV-2000; 2000WO-US30873.

(GETH) GENENTECH INC.

Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W; Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S; Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;

WPI: 2001-408281/43.

P-FSDB: AAU12190.

Isolated, secretory and transmembrane PRO polypeptide used to detect other PRO polypeptides, link bioactive molecules to cells expressing PRO polypeptides, and detect the presence of mammalian tumours e.g. lung, breast, prostate, cervical.

Claim 3: Fig 37: 813pp: English.

CC to factor VIIA. The PRO polypeptides can be used in assays to identify molecules involved in binding interactions. The polynucleotides encoding PRO polypeptides can be used to generate probes, antisense RNA/DNA, transgenic or knock out animals and can be used in gene therapy.

XX SQ Sequence 3501 BP; 921 A; 979 C; 823 G; 776 T; 2 other;

Query Match 26.6%; Score 901; DB 22; Length 3501;

Best Local Similarity 60.0%; Pred. No. 4.8e-226;

Matches 1634; Conservative 0; Mismatches 1062; Indels 27; Gaps 7;

Qy 171 GCTTCTCTCTTCCAGGAGAAAGATTTATCACTTTTGAATATATACCGAATGTATAGCAGTGG 230
 Db 267 GCTTCATCGCTGCAAGAGTCTGAGTACCACCTATGAGTACACGGCGTGTGACAGCACGGG 326
 Qy 231 CTCAGGTGGAGAGTTGCCATTCCAAATTCGACGTGACGTCTGGCCTGCGCTGACCC 290
 Db 327 TTCAGGTGGAGGTGCGCGTACCCGCGCTGTGCACGAGCTGTCTTGAGCC 386
 Qy 291 AGTGAGAGGCAAGAATGCATTTCTCTGTGCTTCTGAGAGTATCTAGAAATCAAGAA 350
 Db 387 CGTCAAGGCGACCGAGTCTCTTCTCTGCAACCCGGGAGTTTCTGGATATGAAGA 446
 Qy 351 CCAGGTATGCAAGTAAAGTGTGAGGACCTATTCTTGGGAGTGGCATCAAAATTTGA 410
 Db 447 CCAGTCATGTAAGCCATCGCGTGAAGGCGCTACTCTCCGCGACAGCAITTCGGTTGA 506
 Qy 411 TGAATGGATGAATTCGCGGAGGATTTCTAACATCGCAACATTCATGGACACATGTGGT 470
 Db 507 TGAGTGGATGAGTGGCGCCCATGGCTTTGCCAGCTCTCAGCCAAACATGGAGCTGATGA 566
 Qy 471 GGGCCCTTCTGACAGCAGCGCGCTGTAACAACTCTTCTTGGATCCCTCGTGGAAA 530
 Db 567 CAGTCTCTGAGTCCA---CCGGAACTGTACTTCGTCGAAGTGGTTCGCCGGGCGCA 623
 Qy 531 CTACATAGAAATCTAATCGTGATGACTGCAGGTGCTTTTGAATCTATGCTGTGCACCTTAA 590
 Db 624 CTACATCGCTCCACACAGGAGCAATGACAGCCACACTGATGTACGGCGTCAACCTGAA 683
 Qy 591 GAAGTCAGGCTATGCTTTTGTAGTACCAGTATCTCGCAACACATCTTCTTTGAGTT 650
 Db 684 GCAATCTGGCAGCGTTAACTTCGAATACTACTATCCAGACTCCAGCATCATCTTTGAGTT 743
 Qy 651 CTTATTCAAAATGATGCTCCAGGAGTGGACACCACTGACAAAGTGGGTAAAAT 710
 Db 744 TTTCTGTCAGAAATGACAGTGCACGCCCAATGCAGATGACTC---CAGGTGATGAAGAC 800
 Qy 711 TACAGCAATGGAAATGGGCTCTCATCTCTGTAATGCTGAATCAGGCACAAACATACT 770
 Db 801 CACAGAGAA---AGGATGGGAATTCACAGTGTGGAGCTAAATCGAGGCAATAATGTCT 857
 Qy 771 CTACTGGAGAACTACAGCATCTTTATGGTTCCTAAGGGGTCAAGCGCTGCTGGTAAA 830
 Db 858 CTTATGGAGAACCAAGCGCTTCTCAGTATGGACCAAAATGACCAAGCTGTGTGGTGG 917
 Qy 831 AAATATCAATTTGAAGGGTGGCGGTACACATACAGAAATGTTTCTTCCAGCCAGGCAC 890
 Db 918 AAACATTCACATAACAGGGTGGCGCTACACTTCAGAAATGCTTCCCTGCAACCTGGCAC 977
 Qy 891 ATTACAGCAAAACAGGTTCATTCAACTGCCAGGTGTGCCAGAAACACCTATTATTCGA 950
 Db 978 GTATGCAGCAAGCAGGAGTCTCTTCTGCAAACTTTGGCCAGCCAACTCTTATTCAA 1037
 Qy 951 GAAAGGACCAAGAAATGTAATAGTGTAAA---CACACTCTCAATTTTCAGGATCCAG 1007
 Db 1038 TAAAGGAAATCTTCTGCCACCAAGTGTGACCTTCACAAATACTCAGAGAAAGATCTTC 1097
 Qy 1008 TGAGTGTACAGAGCGCCCTCCCTGTACCAACAAAGCACTATTTCAGATCCATCTCCATG 1067
 Db 1098 TTCCTGTACGTGCGCCCGAGTTTGCACAGACAAAGATTATTTCACACACACACCGGCTG 1157
 Qy 1068 TGATGAAGAGGAAAGACACAGATAATATACAGTGGATAGAGCCCAAAATCTGCCGGA 1127

Db 1158 CGATGCCAACGGAGAGACACAACATCATGTACAAATGGGCCAAGCCGAAATATCTGTAGCGA 1217
Qy 1128 GGATCTACAGATGCTATTAGATTGCCCTTCTGGAGAGAAGAGGATTTGTCGCCCTTG 1187
Db 1218 GGACTTGAGGGGAGTGAAGTGCCTCTGGTGTGAAGACCCACTGCCACCCTG 1277
Qy 1188 CAACCCCTGGATTTTATACAAATGGATCATCTTCTTGGCATCCCTGTCTCTGCAACAT 1247
Db 1278 CAACCCAGGCTTCTTCAAAACCAACAAAGCAGCCTGCCACCCTGCCCATATGTTCTCTA 1337
Qy 1248 TTCAGATGGACCAAGAATGTAGACCATGTCCAGCAGGAAGGAGCCTGCACCTTGGCTT 1307
Db 1338 TTCCAAATGG---CTCAGATGTACCCCTGCCCCTGCAAGGAGTGAACCTGTGTGGGAT 1394
Qy 1308 TGAATATAAATGGTGGAAATGTCTTCTTCCCTGGCAACATGAAACTTCTCTGCTTCAATGTTGG 1367
Db 1395 TGAATACAATGGTGGACACGCTGCCACAAACATGAAAGACCCGTCTCAGTGGGAT 1454
Qy 1368 GAATTCAAAGTGGATGAATGAATGGTGGAGAGTGGCTGGAGATCATATCCAGAGTGG 1427
Db 1455 CAACCTCGAGTACAAGGATGACAGGCTGGAGGTGGCTGTGATCATCATTTACACAGC 1514
Qy 1428 GGTCTGAGGTTCTGACATGATTAACCTGATTAACCTTGCATATCCAGGATTTAAACC 1487
Db 1515 TGCTGGAGCCTCAGACAATGACTTTCATGATCTCACTCTGTTGTGCCAGGATTTAGACC 1574
Qy 1488 ACCAATCTATGACTGGAGCCAC---GGGTTCTGAACCTAGGAAGAAATGAATTTGTCTT 1544
Db 1575 TCCGAGTCGGTGTGGCAGACACAGAGAAATGAAGGTGGCCAGAAATCACATTTGTCTT 1634
Qy 1545 TGAGACCTCTGTGACGTGACTGTGTTTGTACTTTCATGGTGGATATTAATAGAAAAG 1604
Db 1635 TGAGACCTCTGTGTAACCTGTGAGCTCTACTTTCATGGTGGGTGTAATTTCTAGAC 1694
Qy 1605 TACAATGTGTGTAGAAATCGTGGGTGGAAACCAAGAAAGAAAGCTTACACCATATCAT 1664
Db 1695 CAACACTCTGTGGAGAGTGAAGAGTTCCAAAGGCAACAGTCCATATACCTACATCAT 1754
Qy 1665 CTTCAAGAATGCACTTTTACATTTTACATGGGCATTCAGAGAACTAATCAGGGTCAAGA 1724
Db 1755 TGAGGAAACACTACACAGAGCTTCACTGGGCTTCCAGAGGACCACTTTTTCATGAGGC 1814
Qy 1725 TAATAGACGTTCAATGACATGGTGAAGATTTATTTCTATCAGAGCCACTAATGCACT 1784
Db 1815 AAGCAGGAAGTACACCAATGAGTTGCAAGATCTACTCCATCATGTCAACCAATGTTAT 1874
Qy 1785 TGATGGGTGGCTCTCATGCGCTGCTGCGCTCGGTTCTGAACAGCTCGGTTTCATC 1844
Db 1875 GAATGGGTGGCTCTACTGCGCTCCCTGTGCGCTAGAACCTCTGATGGGCTCCTC 1934
Qy 1845 GTGTGTCCTCCCTCCAGGCCACTACATTTGAGAAAGAAACCAACAGTGCAGGAATG 1904
Db 1935 CTGCACTCTTGTCTGCTGTGTTACTATATTGACCGAGATTCAGGAACCTGCCACTCCTG 1994
Qy 1905 TCCACTGACACTACCTTCCATACATCAGTCTATGGCAAGAGGCTGTGATTTCCATG 1964
Db 1995 CCCCCCTTAAACAAATCTGAAAGCCACAGCCCTATGGTGTCCAGGCTGTGTGCCCTG 2054
Qy 1965 CGGGCTGGGAGTAAAAACAATCAGGACATTCGGTTTGTCTATGACTGCTTTTCTTA 2024
Db 2055 TGGTCCAGGACCAAGAACACAGATCCACTCTGTGCTACATGATTCACCTTCTC 2114
Qy 2025 CCATGAAAGAAAATTCAGATTTTTCATATGACTTTTACCAACCTCAGCAGTGTGGGCTC 2084
Db 2115 AGGCAACACTCCCAACAGGACTTTCAACTTACAACTTCTCCGCTTTGGCAACACCGTCA 2174
Qy 2085 ATTAATGAATGCCCGCCACTTCACTCCAAAGGAAACAAATCTTCCATTTCTCAATAT 2144
Db 2175 TCTTGTGGAGGGCAAGCTTCACTTCCAAAGGTTGAATATCTTCCATTCATCTTACCCT 2234
Qy 2145 CAGTTTATGTGGGATGAGGGAAGAAGATGCTCTCTGTACCAACAAATATTAACAGACTT 2204
Db 2235 CAGTCTCTGTGAACACAGGATAGGAAATGTCTGTGTCCACCGAATGTCTACTGACCT 2294

Qy 2205 TACAGTAAAGAAATAGTGGCAGGTGATGATTTACACAAATTTGTGAGGGCATTTGT 2264
Db 2295 CCGGATTCCTGAGGGTGAGTCAGGG-----TTCTCCAAATCTATCACAGCTACGT 2345
Qy 2265 ATGCCAGTCAACAATATTTCTTCTTGAAGTAAAGGTTTCCGAGCAGCTTATCATCACA 2324
Db 2346 CTGCAGGAGTCAATCATTTCCCCAGGAGGTGACAGGCTTACAGGCCGGGGTTTCTCACA 2405
Qy 2325 ATCATCATCTTGCAGATATCATATAGGAGTCAAGTGTGAAACACATTTGAAATAAT 2384
Db 2406 GCCGTGTCAGCCTTGTGATCACTTATTTGGGTGACAACAGATATGACTCTGGATGGAAT 2465
Qy 2385 TAATATAAAGAAAGATGTTTCCAGTTTCCAACAAGCCAAATACCAGATGTGCAATTTCT 2444
Db 2466 CACCTCCCACTGAACATTTTCCACCTGGAGTCTTGGGAATACCGGACGTGATCTTCTT 2525
Qy 2445 TTATAAGTCTTACAGCAACAACATCTTGTATTAATGGCCGATCAACTGCTGTGAAAT 2504
Db 2526 TTATAGGTTCAATGATGTGACCCAGTCTTGCAGTCTTGGGAGATCAACCACTCCCGGT 2585
Qy 2505 GAGGTGTAATCTACTAATCTGGAGCAGGAGTGAATTTTCACTCCCAAGCAAGTGCACAGC 2564
Db 2586 CAGGTGCACTCCACAGAAACTGTCCCTGGAGTTTGTGCTGCCAGGAAGTGTCTCAGA 2645
Qy 2565 AGGTACCTGTGATGGGTGATGTTCTATTTCTGTGGGAGAGTGTGAAGCTTGGCCCTCT 2624
Db 2646 TGGGACCTGTGATGGCTGCAACTTTCCACTTCTGTGGGAGAGCGGCTGTGTCGCCGT 2705
Qy 2625 GTCTACGAGGAGTGAATCTCCATGAGATTTGAGGAGCCCTGCAAGAGAGATTTTCAGGAAAC 2684
Db 2706 CTGCTCAGTGGCTGACTACCATGCTATCTGTCAGCAGCTGTGTGGTGGGATCCAGAGAC 2765
Qy 2685 CTTGTATCTGTGGAAATGAACCTAATATGTTGCAATTAAGGAATTTCTTTGCCCTGAGAAAA 2744
Db 2766 TACTTACGTGTGNCAGAAACCAAGCTATGCTCTGTGGTGGCATTTCTGCTGACACAG 2825
Qy 2745 GTTGGCAACCTGTGAAACGGTTGACTTTTGGCTGAAGTGGGAGCCGGTGTGGGAGCTTT 2804
Db 2826 AGTCACCATCTGCAAAACCATAGATTTCTGGCTGAAGTGGGCACTCTCTGCAGGCACCTG 2885
Qy 2805 TACTGCCCTTTTGTGGTGGCTCTGACCTGCTTACCTTCTGGAAGAAAGAAATCAAAAGAAA 2864
Db 2886 TACTGCCATCTGCTCACCCTGCTTGGACCTGTGACCTGTACTTTTGGAAAGAAATCAAAACTAGA 2945
Qy 2865 GAAGACCATTTTGAATCTGTTCA 2887
Db 2946 GTACAAGTACTCCAAGCTGGTCA 2968

RESULT 9

AAF89775

ID AAF89775 standard; DNA; 2610 BP.

XX AAF89775;

DT 23-JUL-2001 (first entry)

DE Nucleotide sequence of a human protein expressed in tumour cells.

KW Tumour cell; immunological disease; autoimmune disease; cancer;
infection; ss.

XX Homo sapiens.

FH Key Location/Qualifiers

FT CDS 1..2610

FT /tag= a

FT /product= "tumour expressed protein"

XX WO200131003-A1.

XX 03-MAY-2001.


```
QY 1785 TCATGGGGTGGGCTCTCATGCCCTGCTGTCGCCCTCGTCTCTGAACAGTCGGGTTCAATC 1844
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1620 GAATGGCGTGGCTCTCTACTGCGCTCCCTGTGCCCTAGAACGCTCTGATGTGGGCTCCTC 1679
QY 1845 GTGTGTCCTCCCTCCAGGCCACTACATTTAGAAAGAAACCAACCAAGTGAAGAAATG 1904
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1680 CTGCACTCTTCTGCTGCTGTTACTATATTGACGAGATTGACGAACCTGCACTCCCTG 1739
QY 1905 TCCACTGACACTACCTCTGCTCATCATCAGTGTATGGCAAGAGGCTTGATTCCATG 1964
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1740 CCCCCCTAACAAATCTGAAGCCACAGGCTTATGGTGTCCAGGCTGTGTGCCCTG 1799
QY 1965 CGGGCTGGAGTAAACAAATCAGGACCATTCGGTTGCTATAGTACTGCTTTTTCAT 2024
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1800 TGGTCCAGGACCAAGAACCAAGATCCACTCTCTGTGCTACAAATGATTGCACCTTCTC 1859
QY 2025 CCATGAAAGAAATCAGATTTTTCACATATGACATTTAGCAACCTCAGCAGTGTGGCTC 2084
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1860 AGCAACACTCCCAACGAGGACTTCAACTATCAACTTCTCCCGCTTTGGCAACACCGTCA 1919
QY 2085 ATTAATGAATGGCCCGAGCTTCACCTCCAAAGGAACAAATACTTCCATTTCTTCAATAT 2144
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1920 TCTGCTGAGGCCCAAGCTTCACCTCCAAAGGTTGAATACTTCCATCACTTTACCT 1979
QY 2145 CAGTTATGTGGGATGAGGGAAGAAGATGCTCTCTGTACCAACAAATATTAACAGACTT 2204
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1980 CAGTCTCTGGAAACCAAGGATAGGAAATGCTGTGTGCACCGACAATGTCACTGACCT 2039
QY 2205 TACAGTAAAGAAATAGTGGCAGGTCAGATGATTACACAAATTTGGTAGGGCATTGT 2264
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2040 CCGGATTCCTGAGGTGAGTCAGGGT-----CTCCAATATATCACAGCCTACGT 2090
QY 2265 ATGCCAGTCAACAATATATCTTCTGAAAGTAAAGGTTTCCAGCAGGCTTATCATCACA 2324
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2091 CTGCCAGGAGTCATCATCCCCCAGAGGTGACAGGCTACAGGCGGGGTTTCTCTACA 2150
QY 2325 ATCCATCATCTGGCAGATACATTCATAGGAGTCAGATTTGAACCAACATTTGAAATAT 2384
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2151 GCCTGTGCACTTGTGCTGCACTTATTTGGGTGACACAGCATATGACTCTGGATGGAAT 2210
QY 2385 TAATAAAGAAAGATATCTCCAGTTCCCAAGCAAGCAAAATACAGATGTGCACTTTCT 2444
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2211 CACCTCCAGCTGAATTTTCCACCTGAGTCCCTTGGGAATACCGGAGTGTCTTCT 2270
QY 2445 TTATAAGTCTTACAGCAACAACTTTTGTATTAATGGCCGATCAACTGCTGTGAAAT 2504
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2271 TTATAGTCCAAATGATGTGACCCAGTCTCTGAGTTCTGGGAGATCAACCACTCCGCT 2330
QY 2505 GAGGTGTAATCTACTAAATCTGGAGCAGAGTATTTTCAGTCCCAAGCAAGTGCACG 2564
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2331 CAGGTGCACTCCACAGAAACTGTCCCTGGAAGTTTGTGCTGCCAGAACGTGCTCAGA 2390
QY 2565 AGTACTCTGTGAGTGTACGTTCTTATTTCTGTGGAGAGTGTGAAAGTTCGCCCTCT 2624
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2391 TGGGACCTGTGATGGCTGCACTTCCACTCTGTTGGGAGAGCGGCTGCTGCCGCT 2450
QY 2625 GTGTCGAGAGCATGCTCCATGAGATGAGGAGGCTGCAAGAGAGATTTTTCAGGAAC 2684
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2451 CTGCTCAGTGGGTGACTACCGTGTATCTGACAGCAGTGTGTGGTGGGATCCAGAAGAC 2510
QY 2685 CTTGTATGTGGAATGAACCTTAATGGTGCATTAAGGAATTTCTTCCCTGAGAAAAA 2744
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2511 TACTTACGTGTGGCGAAGACCAAGCTATGCTGCTGGTGGCATTTCTCTGCCCTGAGCAG 2570
QY 2745 GTTGGCAACTGTGAAACGGTTGACTTTTGGCTGAAGGTG 2784
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2571 AGTCACCATCTGCAAAACCATAGATTTCTGGCTGAAAGTG 2610
```

RESULT 10
AAF89774
ID AAF89774 standard; DNA: 2733 BP.
XX

```
AC AAF89774;  
XX 23-JUL-2001 (first entry)  
XX Nucleotide sequence of a human protein expressed in tumour cells.  
DE Tumour cell; immunological disease: autoimmune disease; cancer;  
KW infection; ss.  
XX Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 1..2733  
FT /*tag= a  
FT /product= "tumour expressed protein"  
XX  
PN WO200131003-A1.  
XX  
XX 03-MAY-2001.  
XX 30-OCT-2000; 2000WO-FR03032.  
XX 29-OCT-1999; 99FR-0013629.  
XX (FABR ) FABRE MEDICAMENT SA PIERRE.  
XX Delneste Y, Magistrelli G, Jeannin P, Bonnefoy J;  
XX WPI: 2001-328651/34.  
XX P-PSDB; AAB83850.  
XX  
XX New nucleic acid, expressed in tumours and lymphoid tissue is useful for  
XX identifying agents for treating tumours and autoimmune disease  
XX Claim 2; Page 56-60; 85pp; French.  
XX  
XX The present sequence encodes a human protein expressed in tumour cells.  
XX The polynucleotide is useful for screening cDNA/genomic DNA banks and  
XX for cloning isolated DNA; identifying mutant forms of the gene that  
XX encodes a human protein, where the mutations are associated with  
XX abnormal gene expression, or promoters and regulators of the gene,  
XX particularly for diagnosis; for recombinant expression of the derived  
XX protein; as probes and primers for detection and amplification; and  
XX as antisense therapeutics. The tumour expressed protein is useful for  
XX raising specific antibodies and to screen agents that modulate its  
XX activity, bind to it or interact with it. These agents are potentially  
XX useful for treatment or prevention of diseases associated with abnormal  
XX expression/activity of the protein, particularly immunological diseases  
XX (autoimmune diseases and cancer) or viral, bacterial, fungal or parasitic  
XX infections.  
XX  
SQ Sequence 2733 BP; 656 A; 783 C; 687 G; 607 T; 0 other;
```

Query Match 25.4%; Score 860.8; DB 22; Length 2733;
Best Local Similarity 60.0%; Pred No. 1.6e-215;
Matches 1571; Conservative 0; Mismatches 1022; Indels 27; Gaps 7;

```
QY 171 GCTTCTCTCTCCAGGAGAAAGATTATCATCTTTTGAATATACGGAATGTGATAGCAGTGG 230
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 135 GCTTCATGCTGCAAGAGTCTGAGTACCACATATGATACACGGCTGTGACAGCACGGG 194
QY 231 CTCAGGTGGAGAGTTGCCATTCCAAATTTCTGCAGTGGAGTCTCTGGCTGCTGACCC 290
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 195 TTCCAGGTGGAGGGTGGCGGTGCCCATACCCCGGCGCTGTGCACGAGCTGCTGACCC 254
QY 291 AGTGAGAGGCAAGAAATGCACATTTCTCTCTGCTGCTGAGAGATATCTAGAATTAAGAA 350
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 255 CGTCAAGGCGCCGAGTGTCTCTCTCTGCAACCGCGGGAGTGTCTGATATGAAGA 314
QY 351 CCAGGTATGCAAGTGTGGTGAAGGCAACCTATTTCTTTGGCAGTGGCATCAAAATTGA 410
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 315 CCAGTCATCTAAGCCATGCGCTGAGGGCGGCTACTCCCTCGGCACACAGGCAATTCGGTTTGA 374
```



```
Db 2514 TGGGACCTGTGATGGGTGCAACTTCCACTTCTCTGTGGGAGCGCGGCTGCTTGGCCGCT 2573
QY 2625 GTGTACGGAGCATGACTTCCATGAGATTGAGGGAGCCTGCAAGAGAGGATTTCAGGAAC 2684
Db 2574 CTGCTCAGTGGCTGACTACCGTGCTATGCTGACGAGCTGTGGTGGATCCAGAGAC 2633
QY 2685 CTTGTATGTGTGAATGAACCTAAATGGTGCAATTAAGGAATTTCTTCCCTGAGAAAA 2744
Db 2634 TACTTACGTGTGGCGAAGAACCCCAAGCTATGCTCTGTGGCATTTCTCTGCTGAGCAG 2693
QY 2745 GTTGGCAACCTGTGAACGGTTGACATTTTGGCTGAAGGTG 2784
Db 2694 AGTCACCATCTGCAAAACCATAGATTCTGGCTGAAAGTG 2733

RESULT 11
AAF27997
ID AAF27997 standard; DNA; 2554 BP.
XX
AC AAF27997;
XX
XX 08-MAY-2001 (first entry)
XX Human TR13 receptor coding sequence SEQ ID NO: 1.
XX Human; tumour necrosis factor receptor; TR13; TR14; infection;
KW cancer; autoimmune disease; allergy; inflammatory disease;
KW graft rejection; apoptosis; cardiovascular disease; aneurysm; ds.
XX
XX Homo sapiens.
XX
XX WO200105834-A1.
XX
XX 25-JAN-2001.
XX
XX 14-JUL-2000; 2000WO-US19343.
XX
XX 16-JUL-1999; 99US-0144087.
XX
XX 18-AUG-1999; 99US-0149450.
XX
XX 20-SEP-1999; 99US-0149712.
XX
XX 10-SEP-1999; 99US-0153089.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Ni J, Young PE;
XX
XX WPI; 2001-112682/12.
XX
XX P-PSDB; AAB35328.
XX
XX Nucleic acids encoding 2 human tumor necrosis factor receptor
XX polypeptides ((TR13) and (TR14)), useful for the prevention, diagnosis
XX and treatment of, e.g. cancers, acquired immune deficiency syndrome and
XX hypohidrotic ectodermal dysplasia -
XX
XX Claim 2; Page 366-369; 418pp; English.
XX
XX The present invention provides the protein and coding sequences of the
XX human tumor necrosis factor receptors TR13 and TR14. These sequences are
XX useful in the diagnosis and treatment of many diseases, including cancer,
XX autoimmune diseases, cardiovascular disorders, allergies,
XX neurodegenerative diseases, graft rejection, inflammation, aneurysms and
XX infections.
XX
XX Sequence 2554 BP; 642 A; 722 C; 588 G; 602 T; 0 other;
XX
XX Query Match 20.3%; Score 686.6; DB 22; Length 2554;
XX Best Local Similarity 59.6%; Pred. No. 1e-169;
XX Matches 1256; Conservative 0; Mismatches 834; Indels 19; Gaps 5;
XX
XX 785 CAGGCATCTTATGGTCTTAAGCCGCTCAAGCCCTGCTGTGTAATAAATATACAAATTG 844
Db 18 CAGCCCTTCTCAGTATGGACCAAGTAGTACCACAGCCCTGCTGTGGTGAACAACTTGGCATAA 77
```

```
Db 1154 TTCTGAAGCCACCAGGCTTATGCTGTCAGGCGCTGTGTCCTGTGTCAGGACCA 1213
Qy 1979 AAAACAATCAGACCATCGGTTTCTAGTACTGCTTTTCTTACCATGAAGAAGAA 2038
Db 1214 AGAACAACAGATCCACTCTGTCTCAATGATGTCACCTTTCACGCAACACTCCA 1273
Qy 2039 ATCAGATTTCGACATGACTTTAGCAACCTCAGCAGTGTGGGCTCATTAATGAATGGCC 2098
Db 1274 CCAGGACTTCAACTACACTTCTCCGGTTTGGCAACACCGTCACTCTTGTGGAGGC 1333
Qy 2099 CGAGTTCACCTCCAAAGAACAAATCTTCCATTTCTTCAATATCATGTTTATGTGGC 2158
Db 1334 CAAGCTTCACTTCCAAAGGTTGAATACTTCCATCACTTTTAACTCCAGTCTCTCTGGA 1393
Qy 2159 ATGAGGGAAGAGATGGCTCTCTTACCACAATATACACACTTACAGTAAAGAA 2218
Db 1394 ACCAGGTAGGAATGCTGTGTGACCGACATGTCACTGACCTCCGGATTCCTGAGG 1453
Qy 2219 TAGTGGCAGGTCAGATGATTACAAATTTGGTAGGGCATTTGTATGCCAGTCAACAA 2278
Db 1454 GTGAGTCAGGG-----TTCTCAAATCTATCACAGCCTACGCTGCCAGGCGATCA 1504
Qy 2279 TTATTCCTTCTGAAGTAAGGTTTCCGAGAGCCTTATCATCAAAATCCATTCATTCGG 2338
Db 1505 TCATCCCCCAGAGGTGACAGGCTACAAGCGCGGGTTTCTTCAACAGCGCTGTCAGCCTTG 1564
Qy 2339 CAGATACATTATAGGATGACAGTTCAGACCATGTAACACACATTTGAAATATTAATAAAGAG 2398
Db 1565 CTGATGCACTTATGGGGTGACACAGATATGACTCTGGATGGAATCACTCCCGAGCTG 1624
Qy 2399 ATATGTTCCAGTTCACAAAGCCAAATACCAAGATGTGCTATTTTATAAGTCTTCTTA 2458
Db 1625 AACTTTTCCACCTGGAGTCTTGGGAATACCGGACGTGATCTTTTATAGTCCAAAG 1684
Qy 2459 CAGCAACAACATCTTGTATTAATGGCCGATCAACTGCTGTGAAATGAGGTGTAATCTTA 2518
Db 1685 ATGTGACCCAGCTCTGACGTTCTGGGAGATCAACCACTACCGCGCTCAGGTGCAGTCCAC 1744
Qy 2519 CTAATCTGAGCAGAGTGATTTCAAGTCCCGCAGCAAGTCCCGACAGTACCTGTGATG 2578
Db 1745 AGAAACTGTCCCTGGAGTTTGTGCTGCCAGAACGTGCTCAGATGGGACCTGTGATG 1804
Qy 2579 GGTGACGTTCTATTTCCCTGGGAGTGTGTAAGCTTGCCTCTGTGTAGGAGCATG 2638
Db 1805 GCTGCAACTTCCATCTCTGTGGAGAGCGCGCTTTCGCCGCTCTCTCAGTGGCTG 1864
Qy 2639 ACTTCCATGAGATTGAGGAGCCTGCAAGAGAGGATTTTCAGGAAACCTTGTATGTGGA 2698
Db 1865 ACTACCATGCTATCTGTCAGCAGCTGTGTGGCTGGATCCAGAAGACTACTTACGTGTGGC 1924
Qy 2699 ATCAACCTAAATGGTGCATTAAGGAATTTCTTGCCTCAGAAAGTTGGCAACCTGTG 2758
Db 1925 GAGAACCAAGCTATGCTCTGGTGGATTTCTGCTGAGCAGAGATCACCCTGCA 1984
Qy 2759 AAACGGTTGACCTTTTGGCTGAAGGTGGGAGCGGCTGTGGAGCTTTTACTGCGCTTTTGC 2818
Db 1985 AAACCATAGATTCTGGCTGAAGTGGGCTCTCTCAGGACCTGTACTGCACTCTCC 2044
Qy 2819 TGGTGGCTGTGACCTGCTACTTCTGAAAGAAAGATCAAGAAAGAAAGAGACCAATTTGA 2878
Db 2045 TCACCGCTCTGACCTGCTACTTTTGGAAAAAGAAATCAAAAGACTAGAGTACAAGTACTCA 2104
Qy 2879 ATCTGTTC 2887
Db 2105 AGCTGGTGA 2113
```

RESULT 12

AAX27340

ID AAX27340 standard; DNA; 1129 BP.

XX

AC AAX27340;

```
XX 11-JUN-1999 (first entry)
XX Human secreted protein gene 30 clone H02A26.
DE Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
XX Homo sapiens.
XX WO9902546-A1.
XX 21-JAN-1999.
XX 07-JUL-1998; 98WO-US13684.
XX 12-SEP-1997; 97US-0058785.
XX 08-JUL-1997; 97US-0051916.
XX 08-JUL-1997; 97US-0051918.
XX 08-JUL-1997; 97US-0051919.
XX 08-JUL-1997; 97US-0051920.
XX 08-JUL-1997; 97US-0051925.
XX 08-JUL-1997; 97US-0051926.
XX 08-JUL-1997; 97US-0051928.
XX 08-JUL-1997; 97US-0051929.
XX 08-JUL-1997; 97US-0051930.
XX 08-JUL-1997; 97US-0051931.
XX 08-JUL-1997; 97US-0051932.
XX 08-JUL-1997; 97US-0052732.
XX 08-JUL-1997; 97US-0052733.
XX 08-JUL-1997; 97US-0052793.
XX 08-JUL-1997; 97US-0052795.
XX 08-JUL-1997; 97US-0052803.
XX 18-AUG-1997; 97US-0055684.
XX 18-AUG-1997; 97US-0055722.
XX 18-AUG-1997; 97US-0055723.
XX 18-AUG-1997; 97US-0055947.
XX 18-AUG-1997; 97US-0055948.
XX 18-AUG-1997; 97US-0055949.
XX 18-AUG-1997; 97US-0055950.
XX 18-AUG-1997; 97US-0055953.
XX 18-AUG-1997; 97US-0055954.
XX 18-AUG-1997; 97US-0055964.
XX 18-AUG-1997; 97US-0055984.
XX 18-AUG-1997; 97US-0056360.
XX 12-SEP-1997; 97US-0058660.
XX 12-SEP-1997; 97US-0058661.
XX 12-SEP-1997; 97US-0058664.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Brewer LA, Ebner R, Fischer CL, Kyaw H, Lafleur DM, Li Y, Moore PA;
XX Olsen HS, Rosen CA, Ruben SM, Shi Y, Soppet DR, Zeng Z;
XX WPI: 1999-120770/10.
XX P-PSDB; AAY02679.
XX New isolated human genes and the secreted polypeptides they encode -
XX useful for diagnosis and treatment of e.g. cancers, neurological
XX disorders, immune diseases, inflammation or blood disorders
XX Claim 1; Page 262; 464pp; English.
XX This sequence represents a nucleic acid molecule which encodes a
XX secreted human protein. The gene number, and the clone it is derived
XX from, are detailed in the descriptor line. The gene can be used to
XX generate fusion proteins by linking to the gene to a human immunoglobulin
```

CC Fc portion (e.g. AAX27302) for increasing the stability of the fused
CC protein as compared to the human protein only.
CC The invention relates to 123 novel genes and their fragments (nucleic
CC acid sequences: AAX27311-A2749; amino acid sequences AAY02650-Y02788)
CC which are useful for preventing, treating or ameliorating medical
CC conditions e.g. by protein or gene therapy. Also, pathological
CC conditions can be diagnosed by determining the amount of the new
CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for each of the 123
CC polynucleotides, based on which tissues they are most highly expressed in
CC (see AAX27311 for described uses).

XX SQ Sequence 1129 BP; 379 A; 194 C; 192 G; 363 T; 1 other;

Query Match 14.5%; Score 493.2; DB 20; Length 1129;
Best Local Similarity 98.3%; Pred No. 4.6e-119;
Matches 509; Conservative 0; Mismatches 8; Indels 1; Gaps 1;
QY 2860 AAAAGAAGACCATTGTAATCTGTTCAACTGAAACCTCAAGATCCCAATATATGAA 2919
DB 446 AAAAGAAGACCATTGTAATCTGTTCAACTGAAACCTCAAGATCCCAATATATGAA 505
QY 2920 GAGACAGTCTGTAG-CCTTGAGACTAATGACAAAGAACCTGCTCTAGTTTACAGGA 2978
DB 506 GAGACAGTCTGTAGCCCTTGAGACTAATGACAAAGAACCTGCTCTAGTTTACAGGA 565
QY 2979 CCATATTTTAGGTCGTCTCTACATCTGTCACATTTGGTGATCTCACAGAGGGGCCAT 3038
DB 566 CCATATTTTAGGTCGTCTCTACATCTGTCACATTTGGTGATCTCACAGAGGGGCCAT 625
QY 3039 GCGGCTGAAAGGAGGAGATGGAACATTTGATTCCTTATCATGCTCAAGTACCT 3098
DB 626 GCGGCTGAAAGGAGGAGATGGAACATTTGATTCCTTATCATGCTCAAGTACCT 685
QY 3099 TGCCAATAAAGGAAGCAAGATGTTGGTCTCACTGAAGATGAAGTCAACTCAGCA 3158
DB 686 TGCCAATAAAGGAAGCAAGATGTTGGTCTCACTGAAGATGAAGTCAACTCAGCA 745
QY 3159 AGAGATTTATCTGTATATACATACTGAAACCAAGTTTAAAGCCCAATGCACTGC 3218
DB 746 AGAGATTTATCTGTATATACATACTGAAACCAAGTTTAAAGCCCAATGCACTGC 805
QY 3219 TGATGATCCCATATTAATTTGGTAACTTTTATCTTTATGATGCTACATACAAGT 3278
DB 806 TGATGATCCCATATTAATTTGGTAACTTTTATCTTTATGATGCTACATACAAGT 865
QY 3279 GTGATTTGAAGGCACATGTGAGCATATGCATTATGATCCCAATTTATTTCTTCT 3338
DB 866 GTGATTTGAAGGCACATGTGAGCATATGCATTATGATCCCAATTTATTTCTTCT 925
QY 3339 TTATATTTGGGAAAATTTAAATTTTTTAAAGTAAA 3376
DB 926 TTATATTTGGGAAAATTTAAATTTTTTAAAGTATA 963

RESULT 13
AAS34165
ID AAS34165 standard; cDNA; 672 BP.
XX AC AAS34165;
XX AC AAS34165;
DT 17-DEC-2001 (first entry)
XX Human cDNA encoding a novel foetal antigen, SEQ ID No 689.
DE Human: foetal tissue antigen; ss; antiinflammatory; neuroprotective;
KW immunomodulator; cardiovascular; cytostatic; nephrotoxic;
KW cardiovascular; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; breast neoplasm; cancer;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; food additive.

XX OS Homo sapiens.
XX WO200155312-A2.
XX PD 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US01321.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 18-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.

PR	26-SEP-2000;	2000US-0235484.
PR	27-SEP-2000;	2000US-0235834.
PR	27-SEP-2000;	2000US-0235836.
PR	29-SEP-2000;	2000US-0236327.
PR	29-SEP-2000;	2000US-0236327.
PR	29-SEP-2000;	2000US-0236368.
PR	29-SEP-2000;	2000US-0236369.
PR	29-SEP-2000;	2000US-0236369.
PR	29-SEP-2000;	2000US-0236370.
PR	02-OCT-2000;	2000US-0236802.
PR	02-OCT-2000;	2000US-0237037.
PR	02-OCT-2000;	2000US-0237038.
PR	02-OCT-2000;	2000US-0237039.
PR	02-OCT-2000;	2000US-0237040.
PR	13-OCT-2000;	2000US-0239935.
PR	13-OCT-2000;	2000US-0239937.
PR	13-OCT-2000;	2000US-0239937.
PR	20-OCT-2000;	2000US-0240960.
PR	20-OCT-2000;	2000US-0241221.
PR	20-OCT-2000;	2000US-0241785.
PR	20-OCT-2000;	2000US-0241785.
PR	20-OCT-2000;	2000US-0241787.
PR	20-OCT-2000;	2000US-0241800.
PR	20-OCT-2000;	2000US-0241809.
PR	20-OCT-2000;	2000US-0241826.
PR	01-NOV-2000;	2000US-0244617.
PR	08-NOV-2000;	2000US-0246474.
PR	08-NOV-2000;	2000US-0246475.
PR	08-NOV-2000;	2000US-0246476.
PR	08-NOV-2000;	2000US-0246476.
PR	08-NOV-2000;	2000US-0246477.
PR	08-NOV-2000;	2000US-0246478.
PR	08-NOV-2000;	2000US-0246523.
PR	08-NOV-2000;	2000US-0246524.
PR	08-NOV-2000;	2000US-0246525.
PR	08-NOV-2000;	2000US-0246526.
PR	08-NOV-2000;	2000US-0246527.
PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	08-NOV-2000;	2000US-0246613.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249299.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	06-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	11-DEC-2000;	2000US-0251990.
PR	05-JAN-2001;	2001US-0259678.
XX		

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;
WPI: 2001-488782/53.
P-PSDB: AAU21345.

New polynucleotides and polypeptides for diagnosing, treating, preventing or prognosing e.g. diseases or disorders of the nervous, musculoskeletal, excretory, gastrointestinal, reproductive, and respiratory systems

Claim 1: SEQ ID No 689; 642pp; English.

The invention relates to novel nucleic acids encoding novel human foetal antigens. The nucleic acids and proteins are used to prevent, treat (e.g. by gene therapy) or ameliorate a medical condition in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used in diagnosing a pathological condition or susceptibility to a pathological condition. The antibodies to the antigens can also be used in alleviating symptoms associated with the disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISA). Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors and other nutritional components. Numerous examples of diseases and disorders treated by the nucleic acids and proteins are given in the specification. The present sequence

[illegible]

Db 1272 TCCAAGCTGGTGA 1284

Search completed: May 12, 2003, 00:39:17
Job time : 705.486 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 12, 2003, 00:15:42 ; Search time 181.742 Seconds
(without alignments)
6000.498 Million cell updates/sec

Title: US-10-073-333A-3
Perfect score: 3536
Sequence: 1 atgtgttccgcgcgcggg.....ggtaaaaaaaaaaaaaaaaa 3556

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	494	13.9	1129	4	US-09-227-357-40
2	47.2	1.3	7218	1	US-08-232-463-14
3	43.2	1.2	1120	3	US-09-030-613-1
4	43.2	1.2	1120	4	US-09-451-905-1
5	42	1.2	1209	4	US-09-105-537-21
6	42	1.2	13613	4	US-09-105-537-3
7	42	1.2	38506	3	US-09-320-878-19
8	41.8	1.2	1743	3	US-08-665-259-20
9	41.8	1.2	1743	3	US-08-762-500-20
10	41.8	1.2	1974	3	US-08-762-500-78
11	41.8	1.2	6803	3	US-08-665-259-19
12	41.8	1.2	6803	3	US-08-762-500-19
13	41.2	1.2	2247	2	US-08-524-828-1
14	41.2	1.2	2247	2	US-08-975-114A-1
15	40.8	1.1	16389	4	US-09-741-154-3
16	40.2	1.1	4403765	4	US-09-103-840A-2
17	40.2	1.1	4411529	4	US-09-103-840A-1
18	39.8	1.1	1652	3	US-08-758-662-3
19	39.6	1.1	2834	4	US-09-305-384-6
20	39.6	1.1	6235	4	US-09-305-384-5
21	39.6	1.1	6679	4	US-09-305-384-1
22	39.4	1.1	3507	2	US-08-775-009-36
23	38.8	1.1	9521	4	US-08-972-218-2
24	38.4	1.1	1283	1	US-07-885-970A-17
25	38.4	1.1	1283	1	US-08-298-687A-17
26	38.4	1.1	1283	1	US-08-530-797-18
27	38.4	1.1	1283	1	US-08-298-829-17

RESULT 1
US-09-227-357-40
; Sequence 40, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT APPLICATION NUMBER: US/09/227,357
; CURRENT FILING DATE: 1999-01-08
; EARLIER APPLICATION NUMBER: PCT/US98/13684
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: 60/051,926
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,793
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,925
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,929
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,803
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,732
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,931
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,932
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,916
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,930
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,918
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,920
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,733
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,795
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,919
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,928
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/055,722
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,723
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,948
; EARLIER FILING DATE: 1997-08-18

ALIGNMENTS

28	38.4	1.1	1283	2	US-08-787-335-18	Sequence 18, Appl
29	38.4	1.1	2371	2	US-08-343-443B-1	Sequence 1, Appl
30	38.2	1.1	319	1	US-07-593-657-14	Sequence 14, Appl
31	38.2	1.1	1224	1	US-08-924-847A-1	Sequence 1, Appl
32	38.2	1.1	1224	1	US-08-924-847A-3	Sequence 3, Appl
33	38.2	1.1	1224	3	US-09-120-052-1	Sequence 1, Appl
34	38.2	1.1	1224	3	US-09-120-052-3	Sequence 3, Appl
35	38.2	1.1	1415	4	US-09-345-214-11	Sequence 11, Appl
36	38.2	1.1	1534	1	US-08-300-903A-6	Sequence 6, Appl
37	38.2	1.1	2008	4	US-09-345-214-12	Sequence 12, Appl
38	38.2	1.1	2491	4	US-09-345-214-5	Sequence 5, Appl
39	38.2	1.1	4348	2	US-08-915-868-1	Sequence 1, Appl
40	38	1.1	501	2	US-08-997-080-107	Sequence 107, App
41	38	1.1	501	2	US-08-997-362-107	Sequence 107, App
42	38	1.1	501	4	US-09-095-855-107	Sequence 107, App
43	38	1.1	501	4	US-09-324-543-107	Sequence 107, App
44	38	1.1	501	4	US-09-205-426-107	Sequence 107, App
45	38	1.1	2007	3	US-08-747-221B-36	Sequence 36, Appl

FILE REFERENCE: 600.438US1
CURRENT APPLICATION NUMBER: US/09/105.537A
CURRENT FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 3
LENGTH: 13613
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-105-537-3

Query Match 1.2%; Score 42; DB 4; Length 13613;
Best Local Similarity 62.3%; Pred. No. 0.47;
Matches 66; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
Qy 28 CGGGCAGGGCTGGGGCGGGCGGGAGGCTCCCGGGCGGCGCTCGCGCCCTGG 87
Db 1626 CGGGCGGGCGGTGGAGAGACCTCCGCTAGACCCCGCGGTGCGAGCGCCCGG 1685
Qy 88 AGCCCCGCGCTGGATTGCTGGCGCTCGCGCGCTGCCAGCGG 133
Db 1686 TGGTCCGGGGAGACGAGGCTGGCGGGCGGCGCTCCCGCGCG 1731

RESULT 7
US-09-320-878-19
Sequence 19, Application US/09320878A
Patent No. 6117659
GENERAL INFORMATION:

APPLICANT: ASHLEY, Gary
APPLICANT: BETLACH, Melanie C.
APPLICANT: BETLACH, Mary C.
APPLICANT: MCDANIEL, Robert
APPLICANT: TANG, Li
TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
FILE REFERENCE: 300622002120
CURRENT APPLICATION NUMBER: US/09/320.878A
CURRENT FILING DATE: 1999-05-27
EARLIER APPLICATION NUMBER: CIP OF 09/141.908
EARLIER FILING DATE: 1998-08-28
EARLIER APPLICATION NUMBER: CIP OF 09/073.538
EARLIER FILING DATE: 1998-05-06
EARLIER APPLICATION NUMBER: CIP OF 08/846.247
EARLIER FILING DATE: 1997-04-30
EARLIER APPLICATION NUMBER: 60/119,139
EARLIER FILING DATE: 1998-02-08
EARLIER APPLICATION NUMBER: 60/100,880
EARLIER FILING DATE: 1998-09-22
EARLIER APPLICATION NUMBER: 60/087,080
EARLIER FILING DATE: 1998-05-28
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn ver. 2.0
SEQ ID NO 19
LENGTH: 38506
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-320-878-19

Query Match 1.2%; Score 42; DB 3; Length 38506;
Best Local Similarity 62.3%; Pred. No. 0.95;
Matches 66; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
Qy 28 CGGGCAGGGCTGGGGCGGGCGGGAGGCTCCCGGGCGGCGCTCGCGCCCTGG 87
Db 35623 CGGGCGGGCGGTGGAGAGACCTCCGCTAGACCCCGCGGTGCGAGCGCCCGG 35682
Qy 88 AGCCCCGCGCTGGATTGCTGGCGCTCGCGCGCTGCCAGCGG 133
Db 35683 TGGTCCGGGGAGACGAGGCTGGCGGGCGGCGCTCCCGCGCG 35728

RESULT 8
US-08-665-259-20

Sequence 20, Application US/08665259
Patent No. 6028173
GENERAL INFORMATION:
APPLICANT: Landes, Gregory M.
APPLICANT: Burn, Timothy C.
APPLICANT: Connors, Timothy D.
APPLICANT: Dackowski, William R.
APPLICANT: Van Raay, Terence J.
APPLICANT: Klinger, Katherine W.
TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
COMPOSITIONS, METHODS OF MAKING AND USING SAME
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENZYME CORPORATION
STREET: One Mountain Road
CITY: Framingham
STATE: Massachusetts
COUNTRY: United States of America
ZIP: 01701
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/665.259
FILING DATE: 17-JUN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Dugan, Deborah A.
REGISTRATION NUMBER: 37,315
REFERENCE/DOCKET NUMBER: IG5-9.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 872-8400
TELEFAX: (508) 872-5415
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 1743 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1740
US-08-665-259-20

Query Match 1.2%; Score 41.8; DB 3; Length 1743;
Best Local Similarity 59.8%; Pred. No. 0.14;
Matches 70; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
Qy 6 GTTCCGCGCGCGGGCGGTACGGGGGCGGCTGGGGGGCGGGCGGAGGCTCCCGG 65
Db 1505 GTAGCGCGCTGTGGTCCCGCGGGGATGCGGCTGCGGCTGCGCGCTGCTCCCG 1564
Qy 66 CCGCGGGCGCTCGCGCGCGCTGGAGCGCGCGCTGATTTGCTGGGCGCTCGCGG 122
Db 1565 CCGCGCGCTACCTCTCTGCTGGGGGGCGGCGCTGGAGCGGCGGCTGGGGGG 1621

RESULT 9
US-08-762-500-20
Sequence 20, Application US/08762500
Patent No. 6030806
GENERAL INFORMATION:
APPLICANT: Landes, Gregory M.
APPLICANT: Burn, Timothy C.
APPLICANT: Connors, Timothy D.
APPLICANT: Dackowski, William R.
APPLICANT: Van Raay, Terence J.
APPLICANT: Klinger, Katherine W.
TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
COMPOSITIONS, METHODS OF MAKING AND USING SAME

QY 9 CCGCGCCCGGGGGCGGTACGGGCGAGGGCTGGGGGGGGCGCGGAGGCTCCCGCGCG 68
Db .33 CTGCGCGCGAGCGCGGGATGGGCGAGAGGCTGGCGGCGAGCATCTCCCGCGCG 92
QY 69 CGGGCGCTCGCGCCCTGAGCGCGCGCTGGATTGCTGCTG 110
Db 93 GGGACCCCGGGGGCGCGGAGCGCGCGCGCGCTGCTGCGCG 134

RESULT 14
US-08-975-114A-1
; Sequence 1, Application US/08975114A
; Patent No. 5876714
; GENERAL INFORMATION:
; APPLICANT: Atsushi NISHIKAWA et al.
; TITLE OF INVENTION: HUMAN GLYCOSYLTRANSFERASE GENE,
; TITLE OF INVENTION: COMPOUNDS AND METHOD FOR INHIBITING CANCEROUS METASTASIS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., Suite 800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/975,114A
; FILING DATE: No. 5876714ember 20, 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/107,173
; FILING DATE: August 17, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee Cheng
; REGISTRATION NUMBER: 40,949
; REFERENCE/DOCKET NUMBER: 1-F3439DIV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2247 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
US-08-975-114A-1

Query Match 1.2%; Score 41.2; DB 2; Length 2247;
Best Local Similarity 62.7%; Pred. No. 0.24;
Matches 64; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 9 CCGCGCCCGGGGGCGGTACGGGCGAGGGCTGGGGGGGGCGCGGAGGCTCCCGCGCG 68
Db 33 CTGCGCGCGAGCGCGGGATGGGCGAGAGGCTGGCGGCGAGCATCTCCCGCGCG 92
QY 69 CGGGCGCTCGCGCCCTGAGCGCGCGCTGGATTGCTGCTG 110
Db 93 GGGACCCCGGGGGCGCGGAGCGCGCGCGCGCTGCTGCGCG 134

RESULT 15
US-09-741-154-3/c
; Sequence 3, Application US/09741154
; Patent No. 6437110
; GENERAL INFORMATION:

; APPLICANT: BEASLEY, Ellen M. et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001061
; CURRENT APPLICATION NUMBER: US/09/741,154
; CURRENT FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 16389
; TYPE: DNA
; ORGANISM: Human
US-09-741-154-3

Query Match 1.1%; Score 40.8; DB 4; Length 16389;
Best Local Similarity 58.1%; Pred. No. 1.2;
Matches 72; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 32 GCAGGGGCTGGGGCGCGCGGAGGCTCCCGCGCGGGCGCTCGCGCGCGCTGGAGCC 91
Db 6006 GCGGTCTCTAGCGCGCGCGGCTCTCTCTCGCGCGCGCGCGCGCGCGCGCC 5947
QY 92 CCGCTGGATTGCTGCTGGCGCTCGCGCGCTCGCGCGCGCTGGGACCTGC 151
Db 5946 CTCGCCCGCGCGCTCGCGCGCGCGCGCGCTCGCGCGCGCGCGCGCGCGG 5887
QY 152 CCTC 155
Db 5886 GCGC 5883

Search completed: May 12, 2003, 11:12:52
Job time : 447.742 secs

GenCore version 5.1.5
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 11, 2003, 23:19:12 ; Search time 707.514 Seconds
(without alignments)
11318.654 Million cell updates/sec

Title: US-10-073-333A-3

Perfect score: 3556

Sequence: 1 atgctgttcgcgcgcgggg.....ggtaaaaaaaaaaaaaaaaa 3556

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	3556	100.0	3556	22 AAF75055	TR16-long receptor
2	3214	90.4	3390	22 AAF75054	TR16-short receptor
3	1499.6	42.2	1737	22 AAC84891	Human SEC10 nucleic
4	1182.6	33.3	1508	22 AAC84886	Human SEC5 nucleic
5	991.4	27.9	3280	21 AAA95442	Human CASB619 prot
6	989.8	27.8	3331	22 AAF89765	Nucleotide sequenc
7	987.8	27.8	3501	22 AAS21262	Human cDNA sequenc
8	976.2	27.5	3334	22 AAF28030	Human TR13 recepto
9	860.8	24.2	2610	22 AAF89775	Nucleotide sequenc

10	860.8	24.2	2733	22 AAF89774	Nucleotide sequenc
11	761.4	21.4	2554	22 AAF27997	Human TR13 recepto
12	645.4	18.1	672	22 AAS34165	Human cDNA encodin
13	504	14.2	576	24 ABU69028	Kidney cancer rela
14	494	13.9	1129	20 AAZ27340	Human secreted pro
15	487	13.7	1717	20 AAZ41991	Human endometrium
16	403	11.3	404	22 AAF66437	Novel human polynu
17	381.8	10.7	1149	22 AAF89777	Nucleotide sequenc
18	379.4	10.7	1299	23 ABV23150	Human prostate exp
19	379.4	10.7	1299	23 ABV28991	Human prostate exp
20	285.8	8.0	1119	22 AAH76196	Human seven-transm
21	285.8	8.0	1587	22 AAH76195	Human seven-transm
22	244	6.9	244	22 ABA71400	Human foetal liver
23	244	6.9	244	22 ABA37625	Probe #16091 for g
24	244	6.9	244	22 AAK19720	Human brain expres
25	244	6.9	244	22 AAK45740	Human bone marrow
26	244	6.9	244	22 AAI51665	Probe #20351 used
27	244	6.9	244	22 ABS20017	Human genome-deriv
28	234.6	6.6	627	22 AAF89776	Nucleotide sequenc
29	218	6.1	371	22 ABA58868	Human foetal liver
30	218	6.1	371	22 ABA27759	Probe #6225 for ge
31	218	6.1	371	22 AAK07028	Human brain expres
32	218	6.1	371	22 AAK32769	Human bone marrow
33	218	6.1	371	22 AAI38582	Probe #7268 used t
34	196.4	5.5	457	22 ABS07567	Human genome-deriv
35	196.4	5.5	457	22 AAK32324	Human bone marrow
36	196.4	5.5	457	24 ABS07104	Human genome-deriv
37	196	5.5	196	22 AAK45376	Human bone marrow
38	196	5.5	196	22 ABS19637	Human genome-deriv
39	182.8	5.1	210	22 ABA70282	Human foetal liver
40	182.8	5.1	210	22 ABA36993	Probe #15459 for g
41	182.8	5.1	210	22 AAK18511	Human brain expres
42	182.8	5.1	210	22 AAK44426	Human bone marrow
43	182.8	5.1	210	22 AAI50430	Probe #19106 used
44	182.8	5.1	210	24 ABS18661	Human genome-deriv
45	177.8	5.0	457	22 ABA57660	Human foetal liver

ALIGNMENTS

RESULT 1

AAF75055
ID AAF75055 standard; DNA; 3556 BP.

XX AC AAF75055;

XX DT 10-MAY-2001 (first entry)

XX TR16-long receptor DNA.

XX TR16 receptor: tumour necrosis factor receptor superfamily;
XX apoptosis; inflammatory; cancer; immune; neurodegenerative; ds.

XX Unidentified.

XX WO200112671-A1.

XX PD 22-FEB-2001.

XX 10-AUG-2000; 2000WO-US21885.

XX 12-AUG-1999; 99US-0148348.

XX 13-AUG-1999; 99US-0148683.

XX 13-AUG-1999; 99US-0148870.

XX 16-AUG-1999; 99US-0148758.

XX 17-AUG-1999; 99US-0149181.

XX 18-AUG-1999; 99US-0149453.

XX 19-AUG-1999; 99US-0149498.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Young PE, Baker KP;

XX WPI: 2001-138754/14.
XX New nucleic acid molecule encoding a TR16 tumor necrosis factor
PT receptor polypeptide, useful for the diagnosis and treatment of cancer,
PT autoimmune disorders and cardiovascular diseases -
XX Disclosure: Fig 4; 286pp; English.
XX The present invention relates to a TR16 receptor (tumour necrosis
CC factor receptor superfamily). The invention is useful treating
CC diseases and disorders associated with the inhibited or increased
CC apoptosis. In particular inflammatory diseases, cancers, immune and
CC neurodegenerative disorders may be treated.
XX Sequence 3556 BP; 1043 A; 751 C; 836 G; 926 T; 0 other;
SQ
Query Match 100.08; Score 3556; DB 22; Length 3556;
Best Local Similarity 100.08; Pred. No. 0;
Matches 3556; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCTGTTCCGCGCGCGCGGCTAGCGGCAGGGCTGGGGCGCGCGGAGGCT 60
DB 1 ATGCTGTTCCGCGCGCGCGGCGGCTAGCGGCAGGGCTGGGGCGCGCGGAGGCT 60
QY 61 CCCGCGCGCGGCGCTCGCGCCCTGGAGCCCGCCCTGGATTGCTGGCGCTCGCC 120
DB 61 CCCGCGCGCGGCGCTCGCGCCCTGGAGCCCGCCCTGGATTGCTGGCGCTCGCC 120
QY 121 GCTGCCAGCGCGCTGGGCTGGGACCTGCGCTCTCTCCAGCGCGCGCTTCTCT 180
DB 121 GCTGCCAGCGCGCTGGGCTGGGACCTGCGCTCTCTCCAGCGCGCGCTTCTCTCT 180
QY 181 TGCCAGGAGAAAGATTACATTCAATATACGGAATGTATAGCAGTGGCTCCAGGTGG 240
DB 181 TGCCAGGAGAAAGATTACATTCAATATACGGAATGTATAGCAGTGGCTCCAGGTGG 240
QY 241 AGAGTTGCCATTCCAAATTCGACGTGACCTGCTCTGGCGCTGCGTACCCAGTGAGAGGC 300
DB 241 AGAGTTGCCATTCCAAATTCGACGTGACCTGCTCTGGCGCTGCGTACCCAGTGAGAGGC 300
QY 301 AAAGAAATGCATTTCTCTCTGCTGCTGGAGAGTATCTAGAAATCAAGAACAGGTATGC 360
DB 301 AAAGAAATGCATTTCTCTCTGCTGCTGGAGAGTATCTAGAAATCAAGAACAGGTATGC 360
QY 361 AGTAAGTGTGTGAAGGACCTATTCTTGGGCACTGGCATCAAAATTTGATGAATGGAT 420
DB 361 AGTAAGTGTGTGAAGGACCTATTCTTGGGCACTGGCATCAAAATTTGATGAATGGAT 420
QY 421 GAATTTGCCGGCAGGATTTTCAACATCGCAACATTCATGGACACTGTGGTGGCCCTTCT 480
DB 421 GAATTTGCCGGCAGGATTTTCTAACATCGCAACATTCATGGACACTGTGGTGGCCCTTCT 480
QY 481 GACACAGCGCAGCGGTGAACAACTTCTTCTGATCCCTCGGGAACACTACATAGAA 540
DB 481 GACACAGCGCAGCGGTGTAAACAACTTCTTCTGATCCCTCGGGAACACTACATAGAA 540
QY 541 TCTAATCGTGATGACTGCACCGTGTCTTTGATCTATGCTGTCACCTTAAAGAGTCAGGC 600
DB 541 TCTAATCGTGATGACTGCACCGTGTCTTTGATCTATGCTGTCACCTTAAAGAGTCAGGC 600
QY 601 TATGTCCTTTTGAGTACAGTATGTCACACACATCTTCTTTGAGTCTTTATTCAA 660
DB 601 TATGTCCTTTTGAGTACAGTATGTCACACACATCTTCTTTGAGTCTTTATTCAA 660
QY 661 AATGATCAGTGCAGGAGATGGACACCACTGACAAAGTGGTAAACTTACAGACAAAT 720
DB 661 AATGATCAGTGCAGGAGATGGACACCACTGACAAAGTGGTAAACTTACAGACAAAT 720
QY 721 GGAGAAATGGGCTCTCATTTCTGTAATGCTGAATCAGGCACAAACATCTACTCTGGAGA 780
DB 721 GGAGAAATGGGCTCTCATTTCTGTAATGCTGAATCAGGCACAAACATCTACTCTGGAGA 780

QY 781 ACTACAGGCATCCTTATGGGTTCTAAGCGGTCAGAGCCCTGCTGGTAAAAAATATCACA 840
DB 781 ACTACAGGCATCCTTATGGGTTCTAAGCGGTCAGAGCCCTGCTGGTAAAAAATATCACA 840
QY 841 ATTGAAGGGTGGCGTACACATCAGAAATGTTTCTTCCAGCCAGGCACATTCAGCAAC 900
DB 841 ATTGAAGGGTGGCGTACACATCAGAAATGTTTCTTCCAGCCAGGCACATTCAGCAAC 900
QY 901 AAACCCAGGTTCAATTCACACTGCCAGCTGTCTCCAGAACACACCTATTCTCAGAAAGAGCC 960
DB 901 AAACCCAGGTTCAATTCACACTGCCAGCTGTCTCCAGAACACACCTATTCTCAGAAAGAGCC 960
QY 961 AAAGAATGTATAAGGTGTAAGACCACTCTCAATTTTTCAGGATCCAGTGTACAGAG 1020
DB 961 AAAGAATGTATAAGGTGTAAGACCACTCTCAATTTTTCAGGATCCAGTGTACAGAG 1020
QY 1021 CGCCCTCCTGTACCAAAAAGACTATTTCAGATCCATCTCCTATGTGNTGAAAGGA 1080
DB 1021 CGCCCTCCTGTACCAAAAAGACTATTTCAGATCCATCTCCTATGTGNTGAAAGGA 1080
QY 1081 AAGACACAGATTAATGTACAAAGTGAAGGCCAAATCTGCCGGAGGATCTCACAGAT 1140
DB 1081 AAGACACAGATTAATGTACAAAGTGAAGGCCAAATCTGCCGGAGGATCTCACAGAT 1140
QY 1141 GCTATTAGATTGCCCTTCTGAGAGAAAGAGATTTGTCGCCCTTGCACCCCTGGATTT 1200
DB 1141 GCTATTAGATTGCCCTTCTGAGAGAAAGAGATTTGTCGCCCTTGCACCCCTGGATTT 1200
QY 1201 TATACAATGGATCATCTTCTGCCATCCCTGCTCCTCTGGAACATTTTCAGATGGAACC 1260
DB 1201 TATACAATGGATCATCTTCTGCCATCCCTGCTCCTCTGGAACATTTTCAGATGGAACC 1260
QY 1261 AAAGAATGTAGAACCATGTCCAGCAGGAAGAGCTGTCACCTTGGCTTTGAATATAAATGG 1320
DB 1261 AAAGAATGTAGAACCATGTCCAGCAGGAAGAGCTGTCACCTTGGCTTTGAATATAAATGG 1320
QY 1321 TGGAAATGCTTCTTGGCAACATGAAACTTCCCTTCAATTTGGGAAATTCAAAGTGC 1380
DB 1321 TGGAAATGCTTCTTGGCAACATGAAACTTCCCTTCAATTTGGGAAATTCAAAGTGC 1380
QY 1381 GATGAATCAATGGTTGGAGGTGGCTGGAGATCATATCCAGAGTGGGCTGGAGGTTCT 1440
DB 1381 GATGAATCAATGGTTGGAGGTGGCTGGAGATCATATCCAGAGTGGGCTGGAGGTTCT 1440
QY 1441 GACAAATGATTAACCTTGAATATCCAGAGATTTAAACCCACCAACATCTATG 1500
DB 1441 GACAAATGATTAACCTTGAATATCCAGAGATTTAAACCCACCAACATCTATG 1500
QY 1501 ACTGGAGCCAGGGTTCTGAACATAGGAAGATTAACATTTTGTCTTTGAGACCCCTCTGTTCA 1560
DB 1501 ACTGGAGCCAGGGTTCTGAACATAGGAAGATTAACATTTTGTCTTTGAGACCCCTCTGTTCA 1560
QY 1561 GCTGACTGTGTTTCTACTTCAATGGTGGATATTATAGAAAGTACAAATGTGTTAGAA 1620
DB 1561 GCTGACTGTGTTTCTACTTCAATGGTGGATATTATAGAAAGTACAAATGTGTTAGAA 1620
QY 1621 TCGTGGGTGGAACCAAGAAACAAAGCTTACACCCATATCATCTTCAAGAAATGCAACT 1680
DB 1621 TCGTGGGTGGAACCAAGAAACAAAGCTTACACCCATATCATCTTCAAGAAATGCAACT 1680
QY 1681 TTTACATTTACATGGGCATTCAGAGAACTAATCAGGTCAGATAAATAGACGGTTCAATC 1740
DB 1681 TTTACATTTACATGGGCATTCAGAGAACTAATCAGGTCAGATAAATAGACGGTTCAATC 1740
QY 1741 AATGACATGGTGAAGATTTATTTCTATCAGCCACCTAATGAGGTTGATGGGTGGCGTCC 1800
DB 1741 AATGACATGGTGAAGATTTATTTCTATCAGCCACCTAATGAGGTTGATGGGTGGCGTCC 1800
QY 1801 TCATGCCCTGCTGTGCCCTCGGTTCTGAACAGTGGGTTTCAATCGTGTGTCCTTCCCTT 1860
DB 1801 TCATGCCCTGCTGTGCCCTCGGTTCTGAACAGTGGGTTTCAATCGTGTGTCCTTCCCTT 1860
QY 1861 CCAGGCCACTACATTGAGAAAGAAACCAACCAAGTGCAGGAATGTCCACCTGACACCTAC 1920

```
|||||
Db 1861 CCAGGCCACTACATTGAGAAAGAAACCAACCAAGTGCAGGAATGTCCACCTGACACCTAC 1920
QY 1921 CTGTCCATACATCAGTCTATGGCAAGAGGCTTGTATTCCATCGGGGCTGGGAGTAA 1980
Db 1921 CTGTCCATACATCAGTCTATGGCAAGAGGCTTGTATTCCATCGGGGCTGGGAGTAA 1980
QY 1981 AACAAATCAGGACCATTCGGTCTGTATAGTACTGCTCTTTTCTACCATGAAAGAAAT 2040
Db 1981 AACAAATCAGGACCATTCGGTCTGTATAGTACTGCTCTTTTCTACCATGAAAGAAAT 2040
QY 2041 CAGATTTTGCACATGACTTTTACCAACCTCAGCAGTGTGGGCTCATTAATGAATGCCCC 2100
Db 2041 CAGATTTTGCACATGACTTTTACCAACCTCAGCAGTGTGGGCTCATTAATGAATGCCCC 2100
QY 2101 AGCTTCACCTCCAAAGGAACAAATACTTCCATTCTTCAATATCAGTTATGTGGGCAT 2160
Db 2101 AGCTTCACCTCCAAAGGAACAAATACTTCCATTCTTCAATATCAGTTATGTGGGCAT 2160
QY 2161 GAGGGGAAGAAGATGGCTCTCTGTACCAACAATATACAGACTTTACAGTAAAGAAATA 2220
Db 2161 GAGGGGAAGAAGATGGCTCTCTGTACCAACAATATACAGACTTTACAGTAAAGAAATA 2220
QY 2221 GTGGCAGGTCAGATGATTACACAAATTTGGTAGGGGCATTTGTATGCCAGTCAACAAT 2280
Db 2221 GTGGCAGGTCAGATGATTACACAAATTTGGTAGGGGCATTTGTATGCCAGTCAACAAT 2280
QY 2281 ATTCCCTTCAAGTAGAAGGTTTCCGAGCAGCCTTATCATCACAATCCATCATCTGGCA 2340
Db 2281 ATTCCCTTCAAGTAGAAGGTTTCCGAGCAGCCTTATCATCACAATCCATCATCTGGCA 2340
QY 2341 GATACATTTCATAGGAGTCAGATTGAAACACATTTGAAATAATTAATATAAAGAGAT 2400
Db 2341 GATACATTTCATAGGAGTCAGATTGAAACACATTTGAAATAATTAATATAAAGAGAT 2400
QY 2401 ATGTTCCAGTTCACAAAGCCAAATACCAGATGTGCATTTCTTTTATAGTCTTCTACA 2460
Db 2401 ATGTTCCAGTTCACAAAGCCAAATACCAGATGTGCATTTCTTTTATAGTCTTCTACA 2460
QY 2461 GCAACACATCTTGTATTATGCGCATCAACTCACTGCTGGAATGAGGTCTAATCTCTACT 2520
Db 2461 GCAACACATCTTGTATTATGCGCATCAACTCACTGCTGGAATGAGGTCTAATCTCTACT 2520
QY 2521 AAATCTGGACGAGAGTATTTTCAGTCCCGCAGCAAGTGCAGCAGGACCTGTGTATGGG 2580
Db 2521 AAATCTGGACGAGAGTATTTTCAGTCCCGCAGCAAGTGCAGCAGGACCTGTGTATGGG 2580
QY 2581 TGTACGTTCTATTTCCTGTGGGAGAGTGTGAAGCTTGCCTCTGTGTACGGAGCATGAC 2640
Db 2581 TGTACGTTCTATTTCCTGTGGGAGAGTGTGAAGCTTGCCTCTGTGTACGGAGCATGAC 2640
QY 2641 TTCCATGAGATTGAGGAGCCTGCAAGAGAGGATTTTCAGGAAACCTTGTATGTGGAAT 2700
Db 2641 TTCCATGAGATTGAGGAGCCTGCAAGAGAGGATTTTCAGGAAACCTTGTATGTGGAAT 2700
QY 2701 GAACCTAAATGGTGCATTAAGGAATTTCTTTCCTGAGAAAAAGTTGGCAACCTGTGAA 2760
Db 2701 GAACCTAAATGGTGCATTAAGGAATTTCTTTCCTGAGAAAAAGTTGGCAACCTGTGAA 2760
QY 2761 ACGGTTGACTTTTGGCTGGAAGTGGGAGCGGTGTGGAGCTTTTACTGCCGTTTGGCTG 2820
Db 2761 ACGGTTGACTTTTGGCTGGAAGTGGGAGCGGTGTGGAGCTTTTACTGCCGTTTGGCTG 2820
QY 2821 GTGGCTTGACCTGCTACTTCTGGAAAAAGAAATCAAAAACCTGGAATATATCCAG 2880
Db 2821 GTGGCTTGACCTGCTACTTCTGGAAAAAGAAATCAAAAACCTGGAATATATCCAG 2880
QY 2881 TTAGTAATGACGACTAACTCAAAAGAGTGTGAACCTCCGCTGCAGACAGTGTGTCTATC 2940
Db 2881 TTAGTAATGACGACTAACTCAAAAGAGTGTGAACCTCCGCTGCAGACAGTGTGTCTATC 2940
QY 2941 ATGCAAGGAGAAGATAATGAAGAGGAAAGTTGTATATTCCAAATAAACAGTCACTACTAGGA 3000
|||||
```

```
Db 2941 ATGGAAGGAGAAGATAATGAAGAGGAAGTTGTATATTCCAAATAAACAGTCACTACTAGGA 3000
QY 3001 AAACCTAAATCTTTGCAACCAAGGAAAGAGACCATTTTGAATCTGTTCAACTGAAA 3060
Db 3001 AAACCTAAATCTTTGCAACCAAGGAAAGAGACCATTTTGAATCTGTTCAACTGAAA 3060
QY 3061 ACCTCAAGATCCCAAAATATATGAAGAGACAGTGTGTAGCCTTGAGACTTAATGAACAA 3120
Db 3061 ACCTCAAGATCCCAAAATATATGAAGAGACAGTGTGTAGCCTTGAGACTTAATGAACAA 3120
QY 3121 GAAACCTGCTGTAGTTTACAGGACCATATTTTAGGGTCTGCTCATACCTGTCACT 3180
Db 3121 GAAACCTGCTGTAGTTTACAGGACCATATTTTAGGGTCTGCTCATACCTGTCACT 3180
QY 3181 GGTGATCTCACAGAGAGGCGCATCGCTGAAAGGAGGAGATTGAACATTTTGATT 3240
Db 3181 GGTGATCTCACAGAGAGGCGCATCGCTGAAAGGAGGAGATTGAACATTTTGATT 3240
QY 3241 GCCTTATCACATGGTCAAGTACCTTGCCAAATAAAGGAAAGCAATGATTTGGGTCTCAA 3300
Db 3241 GCCTTATCACATGGTCAAGTACCTTGCCAAATAAAGGAAAGCAATGATTTGGGTCTCAA 3300
QY 3301 CTGAAGATGAAGCTCAACTCAGGAGAGATTTATCTGTATATACACATAACTGAAACCA 3360
Db 3301 CTGAAGATGAAGCTCAACTCAGGAGAGATTTATCTGTATATACACATAACTGAAACCA 3360
QY 3361 AGTTTAAGCCCAATGCACTGCTGATGCATGCCATATTAATTAATGGGTAACCTTTATT 3420
Db 3361 AGTTTAAGCCCAATGCACTGCTGATGCATGCCATATTAATTAATGGGTAACCTTTATT 3420
QY 3421 CTATTATGATCTACATAACAAGTGTGATTTGGAAGGCACATGTGACATATGCATTATG 3480
Db 3421 CTATTATGATCTACATAACAAGTGTGATTTGGAAGGCACATGTGACATATGCATTATG 3480
QY 3481 ATCCAATTAATGTTTTTCTTTCTTTTATATTTTGGGAAAAATTAATAATTTTTTAAGGTA 3540
Db 3481 ATCCAATTAATGTTTTTCTTTCTTTTATATTTTGGGAAAAATTAATAATTTTTTAAGGTA 3540
QY 3541 AAAAAA 3556
Db 3541 AAAAAA 3556

RESULT 2
AAF75054
ID AAF75054 standard; DNA; 3390 BP.
XX
AC AAF75054;
XX
DT 10-MAY-2001 (first entry)
XX
DE TR16-short receptor DNA.
XX
KW TR16 receptor; tumour necrosis factor receptor superfamily;
XX apoptosis; inflammatory; cancer; immune; neurodegenerative; ds.
XX Unidentified.
XX
PN WO200112671-A1.
XX
PD 22-FEB-2001.
XX
PF 10-AUG-2000; 2000WO-US21885.
XX
PR 12-AUG-1999; 99US-0148348.
PR 13-AUG-1999; 99US-0148683.
PR 13-AUG-1999; 99US-0148870.
PR 16-AUG-1999; 99US-0148758.
PR 17-AUG-1999; 99US-0149181.
PR 18-AUG-1999; 99US-0149453.
PR 19-AUG-1999; 99US-0149498.
XX
(PUMA-) HUMAN GENOME SCI INC.
```



```
QY 1861 CCAGGCCACTACATTGAGAAAGAAACCAACCAAGTGCAGGAATGTCCACCTGACACCTAC 1920
DB 1861 CCAGGCCACTACATTGAGAAAGAAACCAACCAAGTGCAGGAATGTCCACCTGACACCTAC 1920
QY 1921 CTGTCCATACATCAGGTCTATGGCAAGAGGCTTGATTTCCATGCGGGCTGGGAGTAA 1980
DB 1921 CTGTCCATACATCAGGTCTATGGCAAGAGGCTTGATTTCCATGCGGGCTGGGAGTAA 1980
QY 1981 AACATCAGGACCATTCGGTTTGCTATAGTACTGCTTTTCTACCATGAAAGAAAT 2040
DB 1981 AACATCAGGACCATTCGGTTTGCTATAGTACTGCTTTTCTACCATGAAAGAAAT 2040
QY 2041 CAGATTTTGCACATGACATTTAGCAACCTCAGCAGTGTGGGCTCATTAATGAATGCCCCC 2100
DB 2041 CAGATTTTGCACATGACATTTAGCAACCTCAGCAGTGTGGGCTCATTAATGAATGCCCCC 2100
QY 2101 AGCTTCACCTCCAAAGGAACAAATACTTCCATTTCTTCAATATACAGTATGTGGGCAT 2160
DB 2101 AGCTTCACCTCCAAAGGAACAAATACTTCCATTTCTTCAATATACAGTATGTGGGCAT 2160
QY 2161 GAGGGAAGAGATGGCTCTCTGTACCAACAAATTTGGTAGGGCATTTGTATGCCAGTCAACAAT 2220
DB 2161 GAGGGAAGAGATGGCTCTCTGTACCAACAAATTTGTATGCCAGTCAACAAT 2220
QY 2221 GTGGCAGGTCAGATGATTACACAAATTTGGTAGGGCATTTGTATGCCAGTCAACAAT 2280
DB 2221 GTGGCAGGTCAGATGATTACACAAATTTGGTAGGGCATTTGTATGCCAGTCAACAAT 2280
QY 2281 ATTCTTCTGAAAGTAAGGGTTTCCGAGCAGCCTTATCATCAATCCATCATCTTCTGGCA 2340
DB 2281 ATTCTTCTGAAAGTAAGGGTTTCCGAGCAGCCTTATCATCAATCCATCATCTTCTGGCA 2340
QY 2341 GATACATTCATAGGAGTCAGATTGAAACCAATTTGAGGAGTGTGATGCAATTAATAAAGAAAT 2400
DB 2341 GATACATTCATAGGAGTCAGATTGAAACCAATTTGAGGAGTGTGATGCAATTAATAAAGAAAT 2400
QY 2401 ATGTTCCCAAGTCCAAACAGCCAAATACCAGATGTGCATTTCTTTTAAAGTCTTCTACA 2460
DB 2401 ATGTTCCCAAGTCCAAACAGCCAAATACCAGATGTGCATTTCTTTTAAAGTCTTCTACA 2460
QY 2461 GCACAAACATCTTGATTAATGGCCGATCACTGCTGTGAAATGAGGTGTAATCTACT 2520
DB 2461 GCACAAACATCTTGATTAATGGCCGATCACTGCTGTGAAATGAGGTGTAATCTACT 2520
QY 2521 AAATCTGGAGCAGGAGTATTTACAGTCCCAAGTGCACAGTGCACAGGTACCTGTGATGG 2580
DB 2521 AAATCTGGAGCAGGAGTATTTACAGTCCCAAGTGCACAGTGCACAGGTACCTGTGATGG 2580
QY 2581 TGTACGTTCTATTTCTGTGGGAGAGTGTGAAGCTTGGCCCTGTGTACGGAGCATGAC 2640
DB 2581 TGTACGTTCTATTTCTGTGGGAGAGTGTGAAGCTTGGCCCTGTGTACGGAGCATGAC 2640
QY 2641 TTCCATGAGATTGAGGAGCCCTGCAAGAGAGATTTTCAGAAACCTTGATGTGGGAAT 2700
DB 2641 TTCCATGAGATTGAGGAGCCCTGCAAGAGAGATTTTCAGAAACCTTGATGTGGGAAT 2700
QY 2701 GAACCTAAATGTTGATTAAGGAATTTCTTTCCTGAGAAAGTTGCAACCTGTGAA 2760
DB 2701 GAACCTAAATGTTGATTAAGGAATTTCTTTCCTGAGAAAGTTGCAACCTGTGAA 2760
QY 2761 ACGTTTGACTTTTGGCTGAAGTGGAGCGGCTGTGGAGCTTTTACTGCCGTTTGTCTG 2820
DB 2761 ACGTTTGACTTTTGGCTGAAGTGGAGCGGCTGTGGAGCTTTTACTGCCGTTTGTCTG 2820
QY 2821 GTGGCTCTGACCTGCTACTTCTGGAAAAAGAAATCAAAA 2880
DB 2821 GTGGCTCTGACCTGCTACTTCTGGAAAAAGAAATCAAAA 2880
QY 2881 TTAGTAATGACGACTAACTCAAAAGAGTGTGAACCTCCGCTCCACAGTGTGCTATC 2940
DB 2859 ----- 2858
```

```
QY 2941 ATGGAAGAGAGAGATTAATGAAGAGGAAGTTGTATATTCCAATAAACAGTCACTACTAGGA 3000
DB 2859 ----- 2858
QY 3001 AAACCTCAAATCTTTGGCAACCAAGGAAAGAAAGACCATTTTGAATCTGTTCAACTGAA 3060
DB 2859 -----GAAAGAAAGACCATTTTGAATCTGTTCAACTGAA 2894
QY 3061 ACCTCAAGATCCCCAATATATCAAGAGACAGTGTGTAGCCTTGAGACTAATGAACAA 3120
DB 2895 ACCTCAAGATCCCCAATATATCAAGAGACAGTGTGTAGCCTTGAGACTAATGAACAA 2954
QY 3121 GAAACCTGCTCTAGTTTACAGGACCATATTTTAGGCTGTGCTCTCATACCTGTCAATT 3180
DB 2955 GAAACCTGCTCTAGTTTACAGGACCATATTTTAGGCTGTGCTCTCATACCTGTCAATT 3014
QY 3181 GGTGATCTCACAGAGAGGGCCATGCCGTGAAAGGAGAGATTTGAAACATTTTGATT 3240
DB 3015 GGTGATCTCACAGAGAGGGCCATGCCGTGAAAGGAGAGATTTGAAACATTTTGATT 3074
QY 3241 GCCTTATCATGTTCAAGTACCTTGCCAAATAAGGAAAGCAATGATTTGGTCTCAA 3300
DB 3075 GCCTTATCATGTTCAAGTACCTTGCCAAATAAGGAAAGCAATGATTTGGTCTCAA 3134
QY 3301 CTGAAGATGAAGCTCAACTCAGGAGAGATTTATCTGTATATACATTAACCTGAAACCA 3360
DB 3135 CTGAAGATGAAGCTCAACTCAGGAGAGATTTATCTGTATATACATTAACCTGAAACCA 3194
QY 3361 AGTTTAAAGCCCAATGCACTGCTGATGCCATATATTAATTTAATGGGTAACCTTTTATT 3420
DB 3195 AGTTTAAAGCCCAATGCACTGCTGATGCCATATATTAATTTAATGGGTAACCTTTTATT 3254
QY 3421 CTTTATGATGTCTACATACAAAGTGTGATTTGGAAGCAGCATGTGAGCATATGATG 3480
DB 3255 CTTTATGATGTCTACATACAAAGTGTGATTTGGAAGCAGCATGTGAGCATATGATG 3314
QY 3481 ATCCAATTTATGTTTTTCTTTCTTTTATATTTTGGGAAATTTAAATTTTTTAAAGGTA 3540
DB 3315 ATCCAATTTATGTTTTTCTTTCTTTTATATTTTGGGAAATTTAAATTTTTTAAAGGTA 3374
QY 3541 AAAAAAAAAAAAAA 3556
DB 3375 AAAAAAAAAAAAAA 3390
RESULT 3
AAC84891
ID AAC84891 standard; cdna; 1737 BP.
XX
AC AAC84891;
XX
DT 20-APR-2001 (first entry)
XX
DE Human SEC10 nucleic acid sequence (clone ID 1795045.0.77).
XX
KW SECX; cytostatic; gynecological; gene therapy; screening assay; human;
KW chromosomal mapping; forensic biology; cell proliferation; cancer;
KW cell differentiation; immune associated disorder; gestational disease;
KW SEC10; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 296..1690
FT /*tag= a
FT /product= "SEC10"
XX
PN WO200078802-A2.
XX
PD 28-DEC-2000.
XX
PF 23-JUN-2000; 2000WO-US17328.
XX
```


RESULT 4

AAC84886
ID AAC84886 standard; cDNA; 1508 BP.

XX AAC84886;

XX 20-APR-2001 (first entry)

XX Human SEC5 nucleic acid sequence (clone ID 1795045.0.61).

XX SEC5; cytostatic; gynecological; gene therapy; screening assay; human;
KW chromosomal mapping; forensic biology; cell proliferation; cancer;
KW cell differentiation; immune associated disorder; gestational disease;
XX SEC5; ss.

XX Homo sapiens.

XX Key Location/Qualifiers
FH 226..1461
FT /*tag= a
FT /*product= "SEC5"

XX WO200078802-A2.

XX 28-DEC-2000.

XX 23-JUN-2000; 2000WO-US17328.

XX 23-JUN-1999; 99US-0140584.

XX 20-JUL-1999; 99US-0144722.

XX 16-SEP-1999; 99US-0154520.

XX 22-JUN-2000; 2000US-0604286.

XX (CURA-) CURAGEN CORP.

XX Shimkets RA, Fernandes E, Vernet C, Yang M, Boldog FL;

PI Herrmann JL;

XX WPI; 2001-071385/08.

XX P-PSDB; AAB48372.

XX Polynucleotides encoding SECX proteins useful for treating disease

PT characterized by an aberrant level of cell proliferation and/or
PT differentiation like cancer or immune associated disorders
XX Claim 3; Fig 5; 132pp; English.
CC The invention relates to human SECX polypeptides and polynucleotides
CC encoding them. The SECX polypeptides can be expressed by standard
CC recombinant methodology. The SECX polypeptides are useful for treating
CC or preventing a SECX-associated disorder. The invention is useful in
CC screening assays; detection assays (e.g. chromosomal mapping, cell and
CC tissue typing, forensic biology); predictive medicine (diagnostic assays,
CC prognostic assays, monitoring clinical trials, and pharmacogenomics); and
CC methods of treatment (e.g. therapeutic and prophylactic); especially
CC disorders characterized by aberrant cell proliferation and/or
CC differentiation like cancer or immune associated disorders or gestational
CC disease. The present sequence represents a SEC5 nucleic acid sequence.

XX Sequence 1508 BP; 444 A; 306 C; 355 G; 402 T; 1 other;

Query Match 33.3%; Score 1182.6; DB 22; Length 1508;
Best Local Similarity 98.4%; Pred. No. 1.8e-300;
Matches 1228; Conservative 0; Mismatches 10; Indels 10; Gaps 3;

QY 506 ACTCTTCTGGATCCCTCGTGAACACTACATAGATCTAATCGTGATGACTGCACGGTGT 565

DB 230 ACTCTTCTGGATCCCTCGTGAACACTACATAGATCTAATCGTGATGACTGCACGGTGT 289

QY 566 CTTTGATCTATCTGTGCACCTTAAGACCTAGGCTATGCTCTTTGAGTACCAGTATG 625

DB 290 CTTTGATCTATCTGTGCACCTTAAGACCTAGGCTATGCTCTTTGAGTACCAGTATG 349

QY 626 TCGACAACAACATCTCTTTGAGTTCTTTATTTCAAATGATCAGTCCAGGAGATGGACA 685
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 350 TCGACAACAACATCTCTTTGAGTTCTTTATTTCAAATGATCAGTCCAGGAGATGGACA 409
QY 686 CCACCACCTGACAAGTGGGTAAACCTTTACAGACAATGGAAATGGGGCTCTCATCTGTATA 745
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 410 CCACCACCTGACAAGTGGGTAAACCTTTACAGACAATGGAAATGGGGCTCTCATCTGTATA 469
QY 746 TCCTGAAATCAGGCACAAACATACCTACTGAGAGAACTACAGGCATCCTTATGGGTTCTA 805
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 470 TGTGAAATCAGGCACAAACATACCTACTGAGAGAACTACAGGCATCCTTATGGGTTCTA 529
QY 806 AGCGGTCAAGCTGTGCTGTTAAATAATACAAATTTGAAGGGTGGCGTACACATCAG 865
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 530 AGCGGTCAAGCTGTGCTGTTAAATAATACAAATTTGAAGGGTGGCGTACACATCAG 589
QY 866 AATGTTTTCTTGCAGCCAGGCACATTCAGCAACAACAGGTTTCATTCACATGCCAGG 925
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 590 AATGTTTTCTTGCAGCCAGGCACATTCAGCAACAACAGGTTTCATTCACATGCCAGG 649
QY 926 TGTGTCCCAAGAACCTATCTGAGAAAGGAGCCAAAGAAATGTAAAGTGTAAAGAGC 985
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 650 TGTGTCCCAAGAACCTATCTGAGAAAGGAGCCAAAGAAATGTAAAGTGTAAAGAGC 709
QY 986 ACTCTCAATTTTC-----AGGATCCAGTGAGTGATACAGAGCCCTCCTGTACCACAA 1039
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 710 ACTCTCAATTTTCAGAGGAGGATCCAGTGAGTGATACAGAGCCCTCCTGTACCACAA 769
QY 1040 AAGACTATTTCCAGATCCATCTCCATGTGATGAAGAGAAAGACACAGATAATGTACA 1099
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 770 AAGACTATTTCCAGATCCATCTCCATGTGATGAAGAGAAAGACACAGATAATGTACA 829
QY 1100 AGTGGATAGAGCCCAAAATCTGCCGGAGGATCTCACAGATGCTATAGATGCCCCCTT 1159
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 830 AGTGGATAGAGCCCAAAATCTGCCGGAGGATCTCACAGATGCTATAGATGCCCCCTT 889
QY 1160 CTGGAGAGAAAGAGGATTCCTGCCCTTGCAACCTCGATTTTATAACAATGGATCATCTT 1219
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 890 CTGGAGAGAAAGAGGATTCCTGCCCTTGCAACCTCGATTTTATAACAATGGATCATCTT 949
QY 1220 CTTGGCATCCCTGTCTCTCGAACAATTTTCAGATGGAACCAAGAAATGTAGACCATGTC 1279
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 950 CTTGGCATCCCTGTCTCTCGAACAATTTTCAGATGGAACCAAGAAATGTAGACCATGTC 1009
QY 1280 CAGCAGGAACGAGCCTGCACCTGGCTTTGAATATAAATGGTGGATGCTCTTCCTGGCA 1339
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 1010 CAGCAGGAACGAGCCTGCACCTGGCTTTGAATATAAATGGTGGATGCTCTTCCTGGCA 1069
QY 1340 ACATGAAACTTCTCTGCTTCAATGTTGGGAATTTCAAAGTCCGATGGAATGATGTTGGG 1399
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 1070 ACATGAAACTTCTCTGCTTCAATGTTGGGAATTTCAAAGTCCGATGGAATGATGTTGGG 1129
QY 1400 AGTGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTGACAATGATACCTGATCT 1459
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 1130 AGTGCTGGAGATCATATCCAGAGTGGGGCTGGAGGTTCTGACAATGATACCTGATCT 1189
QY 1460 TAACTTGATATCCAGATTTAAACCAACCAACATCTATGACTGAGCCACCGGTTCTG 1519
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 1190 TAACTTGATATCCAGATTTAAACCAACCAACATCTATGACTGAGCCACCGGTTCTG 1249
QY 1520 AACTAGAGAATAACATTTGCTTTTGGAGACCTCTGTTGAGTGAATGATGTTTGTACT 1579
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 1250 AACTAGAGAATAACATTTGCTTTTGGAGACCTCTGTTGAGTGAATGATGTTTGTACT 1309
QY 1580 TCATGTTGATATTAATAGAAAAAGTACAAATGTTGTTAGAACTGTTGGGTGGAAACCAAG 1639
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 1310 TCATGTTGATATTAATAGAAAAAGTACAAATGTTGTTAGAACTGTTGGGTGGAAACCAAG 1369
QY 1640 AAAAAACACTTACACCCATATCATCTTCAAGNATCAACTTTTACATTTACATGGG--C 1697
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 1370 AAAAAACACTTACACCCATATCATCTTCAAGNATCAACTTTTACATTTACATGGGCA 1429
QY 1698 ATTCCAGAGAATAAT--CAGGGTCAAGATAATAGACGGTTCATCAAT 1743

Db 1430 TTTCCAGAGAACTAATTCAGGGTCCAGATAAATAGACGGTTCNCCAT 1477

RESULT 5

AAA95442

ID AAA95442 standard: cDNA: 3280 BP.

XX AC AAA95442;

XX DT 12-FEB-2001 (first entry)

XX DE Human CASB619 protein coding sequence #1.

XX KW Human; CASB619; cancer; autoimmune disease; immunogen; vaccine;

XX KW epitope; ss.

XX OS Homo sapiens.

XX FH Key

FT CDS

FT Location/Qualifiers

FT 1..3342

FT /*tag= a

FT /product= "CASB619"

XX PN WO200058460-A2.

XX PD 05-OCT-2000.

XX PF 20-MAR-2000; 2000WO-EP02478.

XX PR 26-MAR-1999; 99GB-0007113.

XX PR 25-SEP-1999; 99GB-0022858.

XX PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX PI Bruck CEM, Cassart J, Coche T, Vinals De Bassols YC;

XX DR WPI: 2000-664923/64.

XX DR P-PSDB: AAB26179.

XX PT Novel CASB619 polypeptides useful for diagnosing, and as vaccines for

XX PT prophylactic and therapeutic treatment of, cancers, particularly

XX PT ovarian and colon carcinoma, and autoimmune diseases

XX PS Claim 13; Page 53-54; 68pp; English.

XX CC The present sequence comprises the human CASB619 coding sequence. This

XX CC protein is thought to be specifically or over-expressed in tumour cells,

XX CC and so can be used as a target for antigen-specific immune responses

XX CC which can cause destruction of the tumour cell. In addition, the protein

XX CC and gene can be used in cancer diagnosis, in the treatment of autoimmune

XX CC diseases and in vaccines against cancer and autoimmune disease. The

XX CC invention provides a number of epitopes derived from the protein which

XX CC can be used as immunogens.

XX SQ Sequence 3280 BP; 810 A; 930 C; 797 G; 743 T; 0 other;

XX

XX Query Match

XX Best Local Similarity 27.9%; Score 991.4; DB 21; Length 3280;

XX Matches 1758; Conservative 0; Mismatches 1116; Indels 27; Gaps 7;

XX

XX Db 171 GCTTCTCCTTGCAGGAGAAAGATTATACACTTTTGAATATAGGAATGTGATGAGCAGTGG 230

XX

XX 135 GCTTCATCGCTGCAAGAGCTGAGTACCACACTAGTACACGGCGTGTGACAGCAGGG 194

XX

XX QY 231 CTCAGGTGGAGATGGCCATTCCAAATTCGACGTGGACTGCTGCGGCTGCTGACCC 290

XX

XX Db 195 TTTCCAGGTGGGGTGGCGGCGGATACCCCGGGCCCTGTGCACAGCTGCTGACCC 254

XX

XX QY 291 AGTGAGAGCAAGATGACATTTCTCTGCTTCTGAGAGATCTAGAAATGAAGA 350

XX

XX Db 255 CGTCAAGGCACCGAGTGTCTCTCTCTGCAACGCCGGGGAGTTTCTGGATATGAAGA 314

QY 351 CCAGGTATGACAGTAAGTGTGGTGAAGGCACCTATTCTTGGCAGTGGCATCAAAATTTGA 410
Db 315 CCAGTCTATGTAAGCCATGCGCTGAGGGCGCTACTCCCTCGGCACAGGCATTCGGTTTGA 374
QY 411 TGAATGGGATGAATTCGGCGCAGGATTTTCTAACATCGCAACATTTCAATGGACACTTGGT 470
Db 375 TGAGTGGGATGAGTGGCCCATGGCTTTGCCAGCCCTTCAGCCAAATGAGAGCTGGATGA 434
QY 471 GGGCCCTTCTGACAGCAGCGCCAGAGCGCTGTAAACAATCTTCTTGGATCCCTCGTGGAA 530
Db 435 CAGTGTCTGAGTCCA---CCGGGAACGTACTTCGTCAAAGTGGGTCCCGCGGGCGGA 491
QY 531 CTACATAGAATCTAATCGTATGACTGCACGGTGTCTTTTGAATCTATCTGTGACCTTAA 590
Db 492 CTACATCGCTCCAAACACGAGCAATGCACGCCACACTGATGTACGCGTCAACCTGAA 551
QY 591 GAAGTCAGGCTATGCTTCTTTGAGTACCAGTATGTGCGACAAACATCTTCTTGTAGTT 650
Db 552 GCAATCTGGCAGCGTTAACTTTCGAATACTATTCAGACTCCAGCATCATCTTTGAGTT 611
QY 651 CTTTATTCAAATGATCAGTGGCCAGGAGATGGACACCACACTGACAAGTGGGTAAAACT 710
Db 612 TTTCTGTCAGATGACCAAGTGCACGCCCAATGCAGATGACTC---CAGGTGGATGAAGAC 668
QY 711 TACAGACAATGGAGAATGGGGCTCTCATTTCTGTAACTCTGAAATCAGGCACAAACATACT 770
Db 669 CACAGAGAA---AGGATGGGAATTCACAGTGTGGAGCTAAATCGAGGCAATATGTCTT 725
QY 771 CTACTGGGAATCAGAGCATCTTATGGTCTTAGGGGGTCAAGCCCTGTGCTGGTAA 830
Db 726 CTATTGGAGAACACACAGCCTTCTCTCAGTATCGACCAAAGTACCCAGCCTGTGCTGAG 785
QY 831 AAATATCACAATTTGAAGGGTGGCGTACACATCAGAAATGTTTCTTGAAGCAGGCAC 890
Db 786 AAACATTCGCAATAACAGGGTGGCCTTACACTTCAGAAATGCTTCCCTGCAAACTGGCAC 845
QY 891 ATTACAGCAACAAACCAGGTTCAATTCACCTGCCAGGTGTGTCCAGAGAAACACTTATTTCTGA 950
Db 846 GTATGCAGACAAGCAGGGCTCTCTTCTGCAAACTTTGCCAGCACAACCTTTATTTCAA 905
QY 951 GAAGCAGGCAAGAAATGTATAGGTGTAA---AGACGACTCTCAATTTTCAGGATCCAG 1007
Db 906 TAAAGGAGAAACTTCTTGGCAGCAGTGTGACCTTGACAAATACTCAGAGAAAGGATTTTC 965
QY 1008 TGAGTGTACAGAGCGCTCCCTGTACACAAAGAACTATTTCCAGATCCATCTCATG 1067
Db 966 TTTCTGTAGTGGCGCCAGCTTGCACAGACAAGATTTATTTCTACACACACAGCGCTG 1025
QY 1068 TGATGAAGAGAAAGAACACACAGATAATGTACAAAGTGGATAGAGCCCAAAATCTGCCGGA 1127
Db 1026 CGATGCCAACGGAGAGACACAACTCATGTACAAATGGGCCAAGCCGAAATCTGTAGCGA 1085
QY 1128 GGATCTACAGATGCTATTAGATTGGCCCTTCTGTGAGAGAGAGAGGATGTCCGCGCTG 1187
Db 1086 GGACCTTGAAGGGGAGTGAAGCTGCTCTGTGTGTAAGACCCACTGCCACCCCTG 1145
QY 1188 CAACCTGTGATTTTATAACAATGGATCATCTTCTTCCATCCCTGCTTCTTCCAGACAT 1247
Db 1146 CAACCTGGCTTCTTCAAAACCAACAAAGCAGCTTCCAGCCCTGCCCATATATGGTTCTTA 1205
QY 1248 TTTCAATGGAACCAAAAGATGTAGACCATGTCCAGCAGGAACGAGCCCTCCACCTTGGCTT 1307
Db 1206 CTCCAATGG---CTCAGACTGTACCCGCTGCCCTGCAGGAGCTGAACCTGCTGTGGAT 1262
QY 1308 TGAATATAATGTTGAATGTCTTCTGCGCAACATGAACATTTCTGCTGCTCAATGTGG 1367
Db 1263 TGAATACAAATGTTGAACACGCTGCCCAACACATGGAACACGCGTCTCTCAGTGGGAT 1322
QY 1368 GAATTCAAAGTGGATGAATGTTGGAGGGTGGCTGGAGATCATATCCAGAGTGG 1427
Db 1323 CAACCTGAGTACAGGGCATGACAGCTGGGAGTGGCTGGTGTGATCAACATTTACACAGC 1382
QY 1428 GGCTGGAGGTCTTGACAATGATTTACCTGATCTTTAAACTTTGCATATCCAGGATTTAAACC 1487

DR WPI: 2001-328651/34.
DR P-PSDB; AAB83845.
XX New nucleic acid, expressed in tumours and lymphoid tissue is useful for
PT identifying agents for treating tumours and autoimmune disease
XX
PS Claim 1; Page 43-48; 85pp; French.
XX
CC The present sequence encodes a human protein expressed in tumour cells.
CC The polynucleotide is useful for screening cDNA/genomic DNA banks and
CC for cloning isolated DNA; identifying mutant forms of the gene that
CC encodes a human protein, where the mutations are associated with
CC abnormal gene expression, or promoters and regulators of the gene,
CC particularly for diagnosis; for recombinant expression of the derived
CC protein; as probes and primers for detection and amplification; and
CC as antisense therapeutics. The tumour expressed protein is useful for
CC raising specific antibodies and to screen agents that modulate its
CC activity, bind to it or interact with it. These agents are potentially
CC useful for treatment or prevention of diseases associated with abnormal
CC expression/activity of the protein, particularly immunological diseases
CC (autoimmune diseases and cancer) or viral, bacterial, fungal or parasitic
CC infections.
XX
XX Sequence 3331 BP; 822 A; 945 C; 813 G; 751 T; 0 other;
SQ
Query Match 27.8%; Score 989.8; DB 22; Length 3331;
Best Local Similarity 60.6%; Pred. No. 1.6e-249;
Matches 1757; Conservative 0; Mismatches 1117; Indels 27; Gaps 7;
QY 171 GCTTCTCTCCAGGAGAAAGATTATACATTTGAAATACGGAATGTGATGAGCAGTGG 230
DB 179 GCTTATGCTGCTCAAGAGTCTGAGTACACATATGATGATACAGGGGTGTGAGCAGCAGGG -238
QY 231 CTCAGGTGGAGATTGCCATTCCAAATTCCTGCAGTGGACTCTCTGGCCCTGCCGTGACCC 290
DB 239 TTCAGGTGGAGGTGCGCGTCCGATACACCGGGCTGTGCACAGCCTGCCGTGACCC 298
QY 291 AGTGAGAGCAAGAAGATCTTCTCTGCTGCTGCTGAGAGATCTAGAAATGAAGAA 350
DB 299 CGTCAAGGCGCCGAGTGTCTCTCTCTCAACCGCGGGAGTTTCTGGATATGAAGA 358
QY 351 CCAGGTATCAGTAAAGTGTGGTGAAGGACCTATTCTTGGGCGAGTGGCATCAAAATTTCA 410
DB 359 CCAGTCATGTAGCCATGCTGAGGGCGCTACTCTCCGACAGGAGATTCGGTTTGA 418
QY 411 TCAATGGGATGAATGCCGGCAGGATTTTCTAACATCCGCAACATTCATGGACACATGTGT 470
DB 419 TGAGTGGGATGAGTGCCTCATGGCTTTGCCAGCCTCTCAGCCAACATGGAGCTGGATGA 478
QY 471 GGGCCCTCTGACAGAGCCGACAGGGCTGTAAACACTCTTCTGGATCCCTCGTGGAAA 530
DB 479 CAGTGTCTGTAGTCCA---CCGGGAACGTGTACTTGTCTCAAGTGGGTTCCCGGGGGCA 535
QY 531 CTACATAGAAATCTAATCGTGAATGCTGACGCGTGTCTTTGATCTATGCTGTGCACCTTAA 590
DB 536 CTACATCGCTCCACACGCGAGATGACAGCCAGCTGATGATGACGCGTCAACCTGAA 595
QY 591 GAAGTCAGGCTATGCTCTCTTTGAGTACAGATGTGCGAACAAACATCTCTTTGAGTT 650
DB 596 GCAATCTGGCCGTTAACTTCCGAATCTACTATCCAGACTCCAGACTCATCTTTGAGTT 655
QY 651 CTTTATTAATGATACAGTCCGAGAGATGGACACCACCACTGACAAAGTGGTAAACT 710
DB 656 TTTCTGTTTCAAGATGACAGTGGCGCCCAATGCAGATGACTC---CAGGTGGATGAAGAC 712
QY 711 TACAGCAATGGAGATGGGGTCTCTATTCTGTGAATGCTGAATCAGGCACAAACATCT 770
DB 713 CACAGAGAA---AGGATGGAAATTCACAGTGTGGAGCTAAATCGAGGCAATATGCTCT 769
QY 771 CTACTGGAGAACTACAGGATCTTATGGGTTCTAAGCGGGTCAAGCCCTGTGCTGTGATAA 830
DB 770 CTATTGGAGAACCAACGCTTCTCAGTATGGACCAAGTACCAGGCTGTGCTGTGTGAG 829

QY 831 AATATACAAATTGAAGGGTGGCGTACATATCAGATGTTTTCCTTGCAGCCAGGCAC 890
DB 830 AAACATTGCCATACAGAGGGTGGCGTACACTTCAAGATGTTTCCCTGCAAAACCTGGCAC 889
QY 891 ATTACGCAACAAACAGGCTTCAATTCAACTGGCAGGTGTGCCAGAAACACATATTCGA 950
DB 890 GTATGACAGACAGCAGGCTCTCTTTCTGCAAACTTTGCCAGCCAACTCTTATTTCAAA 949
QY 951 CAAGAGCCAGCAAGAAATGTAAGGTGTA---AGACGACTCTCAATTTTTCAGGATCCAG 1007
DB 950 TAAAGGAGAAACTTCTGCCACCAGTGTGACCTTCAAAATCTCAGAGMAAGATCTTC 1009
QY 1008 TGAGTGTACAGAGCCCTCTGTTACACAAAGACTATTTTCCAGATCCATATCTCCATG 1067
DB 1010 TTCTCTGTAACGTGCGCCAGCTTGCACAGACAAAGATATTTTCTACACACACACGCGCTG 1069
QY 1068 TCATCAAGAAAGAAACACACAGATATCTACAAAGTGTAGAGCCCAAAATCTGCCGGA 1127
DB 1070 CGATGCCAAGGAGAGACAACTCATGTACAAATTTGGGCGAAGCGGAAATCTGTAGCGA 1129
QY 1128 GGATCTCACAGATGCTATTTAGATTTGCCCTTCTCTGGAGAGAAAGGATTTGTCGCCCTTG 1187
DB 1130 GGACCTTGAGGGGCGAGTGAAGCTGCTGCTGTGTGAAGACCCACTGCCACCCCTG 1189
QY 1188 CAACCTCGATTTTATACAAATGGATCATCTTTCCTGCTCCCTGCTCTCTGGAACATTT 1247
DB 1190 CAACCCAGGCTTCTTCAAAACCAACAGCAGCCTTGCAGCCCTGCCATATATGTCCTTA 1249
QY 1248 TTCAGATGCAACAAAGAAATGTAGACCATGTCCAGCAGAGCGAGCTGTCACTTTGGCTTT 1307
DB 1250 CTCCAATGG---CTCAGACTGTACCCGCTGCGCTGCAGGGACTGAACCTGCTGCTGGATTT 1306
QY 1308 TGAATATAAATGGTGGATTTCTTCTTGGCAACATGAAACTTCTCTGCTTCAATTTGTTGG 1367
DB 1307 TGAATACAAATGCTGGAACACGCTGCCCAACAAACATGGAAGCAGCCGTCTCAGTGGGAT 1366
QY 1368 GAATTCARAAGTCGATGGAATGAATGTTGGAGGTGCTGGAGATCATATCCAGATGG 1427
DB 1367 CAATTCGAGTACAAGGGCATGACAGGTGGAGGTGGGTGATGATACATTTACACAGC 1426
QY 1428 GGCTGGAGTTCTGACAATGATTTAGCTGATCTTAACTTTGCATATCCAGAGATTTAAACC 1487
DB 1427 TGCTGGAGCTTCAGACAAATGACTTTCATGATCTCTACTCTGTTGTGCCAGGATTTAGACC 1486
QY 1488 ACCAATCTATGACTGGAGCCAC---GGTTCCTGAATAGGAAGAAATTAACATTTGCTCTTT 1544
DB 1487 TCCGAGTGGTGTGATGGCAGACAGAAATTAAGAGGTGGCCAGAAATTCATCTCTCTTT 1546
QY 1545 TGAGACCTCTCTGTCAGCTGACTGTTTGTGTTTCTGTTGATGATATTAATAGAAAG 1604
DB 1547 TGAGACCTCTCTGTTGTAAGTGTGAGCTGATGCTTCTGTTGATGTTGTTGTTGTTGTTG 1606
QY 1605 TACAATGTGGTAGAATCGTGGGTGGAAACCAAGAAAGAAAGCTTTACACCCATATCAT 1664
DB 1607 CAACACTCTGTGGAGACGTGGAAGGTTCCTCAAGGCAACAGTCTTATACCTACATCAT 1666
QY 1665 CTTCAAGAAATGCAACTTTTACATTTTACATGGGCTTTCCAGAGAACTAATCAGGGTCAAGA 1724
DB 1667 TGAGGAGAAACACTTACACAGGCTTCCACCTGGCCCTTCCAGAGGAGCCACTTTTTCATGAGGC 1726
QY 1725 TAAATGAGCGTTTCATCAATGACATGTTGAAGATTTATTTCTATCAGAGCCACTTAATGCACT 1784
DB 1727 AAGCAGAAAGTACCAATGAGCTTGGCAAGATCTACTCCATCAATGTCCCAATGTTAT 1786
QY 1785 TGATGGGTGGCTCTCATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1844
DB 1787 GAATGGGTGGCTCTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1846
QY 1845 GTGTGCTCCCTGCCCTCCAGGCGCACTTACATTTGAGAAAGAAACCAACCACTGCAAGGAATG 1904
DB 1847 CTGCACTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1906
QY 1905 TCCACCTGACACCTACCTGCTCCATACATCAGGTCTATGTGCAAGAGGCTGTTGATTTCCATG 1964

Db 1995 CCCCCCTAACAAATTTCTGAAAGCCACACGCGCTTATGTGTCCAGCGCTGTGTGCCCTG 2054
QY 1965 CGGGCTGGAGTAAAAACAATCAGGACCATTCGGTTTGTCTAGTACTGCTTTTCTA 2024
Db 2055 TGGTCCAGGACCAAGAACAAAGATTCACCTCTGTGCTACATGATGACCTTCTC 2114
QY 2025 CCATGAAAAAAGAAATCAGATTTTGCACATGACTTTAGCAACCTCAGCAGTGTGGGCTC 2084
Db 2115 AGCAACACTCAACACGAGCTTCACTACACTTCTCCGCTTTGGCAACACACCGTCA 2174
QY 2085 ATTAATGATGCCCGCAGCTTCACCTCCAAAGGAACAATACTTCCATTTTCAATAT 2144
Db 2175 TCTGTGAGGCGCCAGCTTCACCTTCCAAAGGTTGNAATACTTCCATCACTTTACCT 2234
QY 2145 CAGTTATGTGGCATGAGGGGAAGATGGCTCTGTACCAACAATATACACACTT 2204
Db 2235 CAGTCTCTGGNAACACAGGTAGGAATGTGTGTCCACCGACATGCTACTGACCT 2294
QY 2205 TACAGTAAAGAAATAGTGGCAGGTCAGATGATTACACAAATTTGGTAGGGCATTGT 2264
Db 2295 CGGGATTCTGTAGGGTGAGTCAGG-----TTCTCCAAATCTATCACAGCCTACGT 2345
QY 2265 ATGCCAGTCAACAATTTATTCCTTCTCAAGTAGGGTTTCCGAGCAGCTTTATCATCACA 2324
Db 2346 CTGCCAGGCGATCATATCCCCCAGAGGTGACAGGCTACAAGGCGGGGTTTCCCTACA 2405
QY 2325 ATCCATCTTCTGGCAGATACATTTAGGAGTCACAGTTGAAACACATTTGAAAAATAT 2384
Db 2406 GCCTGTACGCTTGTGATCGACTTATTTGGGTGACACAGATATGACTCTGGATGGAAT 2465
QY 2385 TAATATAAAGAAATATGTTCCAGTTCCTCAACAAAGCAATACCAAGATGTGCAATTTCT 2444
Db 2466 CACCTCCCGAGCTGAATTTTCCACCTGGAGTCTTGGGAATACCGGACGTGATCTTCT 2525
QY 2445 TTATAGTCTTCTACAGCAACACATCTTGTATTAATGGCCGATCAACTGCTGTGAAAT 2504
Db 2526 TTATAGTCCAAATGATGTGACCCAGTCTGTCAGTTCTGGAGATCAACCACTCCGCGT 2585
QY 2505 GAGGTGAATCTCTACTAAATCTGGAGCAGGAGTGATTTTCAGTCCCGAGCAAGTCCGAGC 2564
Db 2586 CAGGTGCACTCCACAGAAACTGTCCCTGGAAGTTGCTGTGCCAGGAACGTGCTCAGA 2645
QY 2565 AGTACCTGTGAGGTGTACGTTCTTCTGTGGGAGAGTGTGAAAGCTTGGCCCTCT 2624
Db 2646 TGGGACCTGTGATGGTGTCAACTTCCACTTCCCTGTGGGAGCGCGCTTGGCCGCT 2705
QY 2625 GTGTAGGAGCATGACTTCCATGAGATTCAGGAGCCTGCAAGAGAGATTTCAAGAAAC 2684
Db 2706 CTGCTCAGTGGCTGACTACCATGCTATCTGTCAGCAGCTGTGTGGCTGGATCCAGANGAC 2765
QY 2685 CTGTATGTGTGGAATGAACCTAAATGGTGCAATTAAGGAATTTCTTTCCTCGAGAAAAA 2744
Db 2766 TACTTACGTGTGNCGASAAACCCAGCTATGCTGTGTGGCATTTCTGCTGAGCAGAG 2825
QY 2745 GTTGGCAACTGTGAAACGGTTGACTTTTGGCTGAAGGTGGGAGCCGCTGTGGAGCTTT 2804
Db 2826 AGTCACCATCTGCAAAACCATAGATTTCTGGCTGAAAGTGGGCATCTCTGCGAGCACCTG 2885
QY 2805 TACTGCGGTTTGTGCTGGTCTGACCTGCTACTTCTGGAAGAAAGTCAAAACCTGGA 2864
Db 2886 TACTGCCATCTCTGCTCACCGCTCTGACCTGCTACTTTTGGAAAAAGAAATCAAAACATGA 2945
QY 2865 ATACAAATATTCACAGTTAGTAATGACGACTAACTCAAAAGAGCTGTGAATCCCGGCTGC 2924
Db 2946 GTACAAGTACTCCAAGCTGGTGTGATGCTACTCTCAAGGAGCTGTGACCTGCCAGCAGC 3005
QY 2925 AGACAGTTGTGCTATCATGGAAGGAGAAGATAATTAAGAGGAAGTTGTATATTCCAAATA 2984
Db 3006 TGACAGCTGCGCCATCATGGAAGCGAGGATGTAGAGGACGACCTCATCTTTTACCAGCAA 3065
QY 2985 ACAGTCACTACTAGGAAACTCAATCTTTTGGCAACCAAGGAAAAAGACACCATTTTGA 3044
Db 3066 GAAGTCACTTTTGGGAAGATCAATCATTTTACCTCCAAGAGGACTCTCTGATGGATTGA 3125

QY 3045 ATCTGTTCAACTGAAACCTC 3065
Db 3126 CTCAGTCCCGCTGAAGACATC 3146

RESULT 8

AAF28030
ID AAF28030 standard; DNA; 3334 BP.

XX AAF28030;

XX 08-MAY-2001 (first entry)

XX Human TR13 receptor coding sequence SEQ ID NO: 39.

XX Human; tumour necrosis factor receptor; TR13; TR14; infection;

KW cancer; autoimmune disease; allergy; inflammatory disease;

KW graft rejection; apoptosis; cardiovascular disease; aneurysm; ds.

XX Homo sapiens.

XX WO200105834-A1.

XX 25-JAN-2001.

XX 14-JUL-2000; 2000WO-US19343.

XX 16-JUL-1999; 99US-0144087.

PR 18-AUG-1999; 99US-0149450.

PR 20-SEP-1999; 99US-0149712.

PR 10-SEP-1999; 99US-0153089.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Ni J, Young PE;

XX WPI; 2001-112682/12.

XX Nucleic acids encoding 2 human tumor necrosis factor receptor

PT polypeptides (TR13) and (TR14)), useful for the prevention, diagnosis

PT and treatment of, e.g. cancers, acquired immune deficiency syndrome and

PT hypohidrotic ectodermal dysplasia

XX Claim 4; Page 394-398; 418pp; English.

XX The present invention provides the protein and coding sequences of the

CC human tumour necrosis factor receptors TR13 and TR14. These sequences are

CC useful in the diagnosis and treatment of many diseases, including cancer,

CC autoimmune diseases, cardiovascular disorders, allergies,

CC neurodegenerative diseases, graft rejection, inflammation, aneurysms and

CC infections.

SQ Sequence 3334 BP; 820 A; 952 C; 811 G; 751 T; 0 other;

Query Match 27.5%; Score 976.2; DB 22; Length 3334;

Best Local Similarity 60.5%; Pred. No. 6.1e-246;

Matches 1755; Conservative 0; Mismatches 1118; Indels 28; Gaps 8;

QY 171 GCTTCTCTCTTGGCAGGAAAGATTATCACTTTTGAATATACGGAATGTGATAGCAGTGG 230
Db 192 GCTTCAGCGCTGCAAGAGCTGTGAGTACCCTATGAGTACACGCGGTGTGACAGCACGGG 251
QY 231 CTCAGGTGAGAGTGTCCATTTCCAAATTCCTGAGTGGAGTGTCTGGCTGCTGACCC 290
Db 252 TTCCAGGTGGAGGTGCGCGTCCGCGATACCCGCGCTGTGCACACGCTGCTGACCC 311
QY 291 AGTGAGAGGCAAGAAATGCACTTTCTCTGCTGCTTCTGGAGAGTATCTAGAAATGAAGAA 350
Db 312 CGTCAAGGCGCAGGTGCTCTTCTCTCAACGCGCGGAGTTCTTGGATATGAAGGA 371
QY 351 CCAGGTATGCAATGAGTGTGTGTAAGGACCTATTTCTTGGGAGTGGCATCAATTTGA 410

Db 372 CAGTCATGTAAGCCATGCCCTGAGGGCCCTACTCCCTCGGCACAGGCAATTCGGTTTGA 431
QY 411 TGAATGGGATGAATTTGCCGACGAGATTTTCTAAACATCGCAACATTTATGACACACTGTGGT 470
Db 432 TGAGTGGGATGAGCTGCCCATGCTTTGCCAGCCTCTCAGCCAAACATGGAGCTGGATGA 491
QY 471 GGGCCCTTCTGACAGAGCCGACAGGCTGTACACACTCTTCTTGGATCCCTCGTGAAA 530
Db 492 CAGTGTCTGTAGTCCA--CCGGGAACCTGACTTCTGCTCAAGTGGGTTCGCCGGGCGA 548
QY 531 CTACATAGAACTTAATCGTGACATGCACGGTCTCTTTGATCTATGCTGACACCTTAA 590
Db 549 CTACATGCCCTTCAACACGAGCAATGCACAGCCACACTGATGAGCCGCTCAACCTGAA 608
QY 591 GAAGTCAGGCTATGCTCTTTTGGTACCACTAGTGTGCAACAAACACTCTTTTGGATT 650
Db 609 GCAATCTGGCACCGTTAACTTCGAATACTACTATCCAGACTCCAGCATCATCTTGGATT 668
QY 651 CTTTATTCAAANTGATCAGTGGCCAGAGATGGACACACCACTGACAAAGTGGGTAAACT 710
Db 669 TTTCTGTTTCAAAATGACAGTGGCCAGCCCAATGCAGATGACTC--CAGGTGGATGAAGAC 725
QY 711 TACAGACAATGGAGATGGGGCTCTCATTTCTGTAANTGCTGAATCAGGCACAAACATACT 770
Db 726 CACAGAAA--AGATGGGAATTCACAGTGTGGAGCTAAATCGAGGCAATTAATGTCT 782
QY 771 CTACTGGAGAACTACAGGCATCTTATGGTCTTAAGCGGTCAAGCCTGTGCTGTGATAA 830
Db 783 CTATGGAGAACCAACAGCCTTCTCAGTATGGACCAAAAGTACCAAGCCTGTGCTGTGAG 842
QY 831 AAATATCACAAATTTGAAGGGTGGGTACACATCAGAAATGTTTCTTTCGAAAGCCAGGCAC 890
Db 843 AAACATTTGCCATAAACAGGGTGGCTACACTTCAGAAATGCTTCCCTGCAAAAGCTGGCAC 902
QY 891 ATTTCAGCAACCAACAGGTTCAATTAACCTGCCAGGTGTGTCAGAAACACCTATTCTGA 950
Db 903 GTATGCAGACACAGGGCTCTCTTCTGCAAACTTTGCCAGCCCAACTCTTATTCAAA 962
QY 951 GAAAGGAGCAAAAGATGTATAAGGTAA--AGACGACTCTCAATTTTCAGGATCCAG 1007
Db 963 TAAAGGAGAAACTTCTTGCCACCACTGTACCCCTGACCAAAATCTCAGAGAAAGATCTTC 1022
QY 1008 TGAGTGTACAGAGCCCTCCCTGTACCACAAAGACTATTTCAGATCCATCCATG 1067
Db 1023 TCTCTGTAAAGCTGGGCCAGCTTGCACAGACAAGATATTCTACACACACAGCCCTG 1082
QY 1068 TGATGAGAAAGAAAGACAGATAATGTACAAAGTGGATAGAGCCCAAAATCTGCCGGGA 1127
Db 1083 CGATGCCAACGGAGACACAACTCATGTACAAATGGCCCAAGCCGNAATCTGTAGCGA 1142
QY 1128 GGATCTCAGAGTGTATTAGATTGGCCCTTCTGGAGAGAAAGATTTGCCGCTTG 1187
Db 1143 GGACCTTGAGGGGCGAGTGAAGCTGCTGCTGTGTGGAAGACCCACTGCCACCCCTG 1202
QY 1188 CAACCTTGGATTTTATAACAAATGGATCATCTTCTGGCATCCCTGTCCCTCGAACAAT 1247
Db 1203 CAACCCAGGCTTCTTCAAAACCAACACAGCACTGCCAGCCCTGCCCATATGTTCCCTA 1262
QY 1248 TTCAGATGGAACCAAGATGTAGACCATGTCCAGCAGGAACGAGCCCTGCACCTTGGCTT 1307
Db 1263 CTCAATGG--CTCAGACTGTACCCGCTGCCCTGCAGGACTGAACCTGCTGGGATT 1319
QY 1308 TGAATATAAATGGTGAATGCTCTTCTGGCAACATCAAAACTTCTGCTTCAATGTGG 1367
Db 1320 TGAATACAAATGGTGAACACGCTGCCCAACAAACATGGAAACGACCTTCTCAGTGGGAT 1379
QY 1368 GAATTCNAAGTCGGATGGATGAATGGTTGGGAGGTGGCTGGAGATCATATCCAGAGTGG 1427
Db 1380 CAACCTTCGAGTACAAAGGCAATGACAGGCTGGGAGGTGGCTGGTGAATCACAATTTACACAGC 1439
QY 1428 GGCTGAGGTTCTGACAATGATTTACCTGATCTTAACTTGCATATCCACGATTTAAACC 1487
Db 1440 TGCTGGAGCCTCAGACAAATGACTTCATGATCTCAGTCTGGTGTGCGCAGGATTTAGACC 1499

QY 1488 ACCAACATCTATGACTGGAGCCAC--GGGTCTGAACTAGGAAGAAATAACATTTGTCTT 1544
Db 1500 TCCGCACTCGGTGATGGCACACAGAGAAATAAAGAGTGGCCAGAAATACATTTGTCTT 1559
QY 1545 TGAGACCCCTCTGTTTACGTGACTGTGTGTTTGTACTTCTATCTGTTGGATATTAATAGAAAAG 1604
Db 1560 TGAGACCCCTCTGTTTCTGTAACCTGTGAGCTCTACTTCTCATGGTGGTGAATCTTAGGAC 1619
QY 1605 TACAAATGTGTAGAAATCGTGGGTGCAACCAAGAAAACAAAGCTTTACACCCATATCAT 1664
Db 1620 CAACACTCTGTGGAGACGTGGAAAGGTTCCAAAGGCAACAGTCCCTATACCTACATCAT 1679
QY 1665 CTTTCAAGAATGCAAACTTTTACATTTTACATGGCATTTCCAGAGAACTAANTCAGGCTCAAGA 1724
Db 1680 TGAGGAGAACACTTACCACAGAGCTTCCACTGGGCTTCCAGAGGACCACTTTTCTGAGGC 1739
QY 1725 TAAATAGACGTTTCAATCAATGACATGGTGAAGATTTATTTCTATCACAGCCACTAANTGACGT 1784
Db 1740 AAGCAGGAAGTACACCAATGACCTTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1799
QY 1785 TGATGGGTGGGCTTCTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1844
Db 1800 GAATGGGTGGGCTTCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1859
QY 1845 GTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1904
Db 1860 CTGCACTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1919
QY 1905 TCCACTTGACACCTTACTTGTCCATACATCAGGTCTATGGCAAAAGAGGCTTGTATTTCCATG 1964
Db 1920 CCGCCCTTAACACAAATTTCTGAAAGCCCAACCACTTATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1979
QY 1965 CGGGCTTGGGAGTAAACAAACAGGACCACTTGGTGTGCTATAGTACGCTTTTCTTCTA 2024
Db 1980 TGGTCCAGGAGCCACAGAACCAAGATCCACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2039
QY 2025 CCATGAAAAAAGAAATCAGATTTTGCATATGACTTTAGCAACCTCAGCAGTGTGGGCTC 2084
Db 2040 ACACAACACTTCAACAGGACTTTCAACTACAACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2099
QY 2085 ATTAAGTAATGGCCCACTTCCACTTCCCAAGGAGGCTTCCAAAGGCTTCCAAAGGCTTCCAAAGGCT 2144
Db 2100 TCTTGTGGGAGGCTTCCACTTCCAAAGGCTTCCAAAGGCTTCCAAAGGCTTCCAAAGGCTTCCAAAGGCT 2159
QY 2145 CAGTTTATGTGGGCTGAGGAGGAGAGTGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2204
Db 2160 CAGTCTCTGTGGAAACCAAGGCTAGGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2219
QY 2205 TACAGTAAAGAAATAGTGGCAGGCTCAGATGATACACAAATTTGGTAGGGGCTTTGT 2264
Db 2220 CCGGATTCCTTGGGCTGAGTCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2270
QY 2265 ATGCCAGTCAACAAATTTATCTCTGTAAGTAAAGGCTTCCGAGCAGGCTTATCATCACA 2324
Db 2271 CTGCCAGGCT 2330
QY 2325 ATCCATCTTCTGGCAGATACATTTATAGGAGTACAGGCTTGAACCAACATTTGAAACAAAT 2384
Db 2331 GCCTGTGAGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2390
QY 2385 TAAATATAAAGAGATATGTTTCCAGTTCACCAAGGCTTCCAGGCTTCCAGGCTTCCAGGCTTCCAGGCT 2444
Db 2391 CACCTTCCCAAGCTTCTTCCAGCTTGGAGTCTTGGGAAATACCAGGCTTCCAGGCTTCCAGGCTTCCAGGCT 2450
QY 2445 TTATAAGTCTTCTACAGCACAACATCTTGTATTAATGGCCGATCAACTGCTGCTGCTGCTGCTGCTGCTGCT 2504
Db 2451 TTATAGTCCCAATGATGTGACCAAGCTTCCAGGCTTCCAGGCTTCCAGGCTTCCAGGCTTCCAGGCTTCCAGGCT 2510
QY 2505 GAGTGTAACTCT 2564
Db 2511 CAGGTGCACTCCACAGAAACTGTCCCTGGAAGTTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2570


```
FT /*tag= a
XX /product= "tumour expressed protein"
PN WO200131003-A1.
XX
XX 03-MAY-2001.
PD
XX
PF 30-OCT-2000; 2000WO-FR03032.
XX
PR 29-OCT-1999; 99FR-0013629.
XX
XX (FABR ) FABRE MEDICAMENT SA PIERRE.
PA
XX Delneste Y; Magistrelli G, Jeannin P, Bonnefoy J;
XX WPI; 2001-328651/34.
XX DR P-PSDB; AAB83850.
XX
XX New nucleic acid, expressed in tumours and lymphoid tissue is useful for
PT identifying agents for treating tumours and autoimmune disease
XX
XX Claim 2; Page 56-60; 85pp; French.
XX
CC The present sequence encodes a human protein expressed in tumour cells.
CC The polynucleotide is useful for screening cDNA/genomic DNA banks and
CC for cloning isolated DNA; identifying mutant forms of the gene that
CC encodes a human protein, where the mutations are associated with
CC abnormal gene expression, or promoters and regulators of the gene,
CC particularly for diagnosis; for recombinant expression of the derived
CC protein; as probes and primers for detection and amplification; and
CC as antisense therapeutics. The tumour expressed protein is useful for
CC raising specific antibodies and to screen agents that modulate its
CC activity, bind to it or interact with it. These agents are potentially
CC useful for treatment or prevention of diseases associated with abnormal
CC expression/activity of the protein, particularly immunological diseases
CC (autoimmune diseases and cancer) or viral, bacterial, fungal or parasitic
XX infections.
XX
XX Sequence 2733 BP; 656 A; 783 C; 687 G; 607 T; 0 other:
XX
XX Query Match 24.2%; Score 860.8; DB 22; Length 2733;
XX Best Local Similarity 60.0%; Pred. No. 1.3e-215;
XX Matches 1571; Conservative 0; Mismatches 1022; Indels 27; Gaps 7;
XX
QY 171 GCTTCCTCTCCAGGAGAAAGATTATACATTTTGAATATACGGAATGTGATAGCAGTGG 230
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
135 GCTTCATGCTCAAAGAGCTGTGATACACTATGATACACGGCGTGTGACAGCACGG 194
QY 231 CTCAGGTGGAGAGTTGCCATTCCTCAAAATTCGACGTGGACTGCTCTGGCCTGCCAGCCC 290
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
195 TTCCAGGTGGAGGGTGGCGGTGCCGATACCCCGGGCCTGTGCACAGCGCTGCCCTGACCC 254
QY 291 AGTGAGAGCAAGAAGATGACATTTCTCTGTGCTTCTGGAGAGTATCTAGAATGAAGAA 350
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
255 CGTCAAGGCCACCGAGTGTCTCTCTCTGCAACCGCGGGAGTTCTGGATATGAAGA 314
QY 351 CCAGGTATGACATAGTGTGTGAAGGCACCTATTCTCTTGGCAGTGGCATCAAAATTTGA 410
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
315 CCAGTCATGTAAGCCATCGCTGAGCGCGCTACTCTCCCTCGGCACAGGATTCGGTTGA 374
QY 411 TGAATGGGATGAATTTGCCGGCAGGATTTTCTAACATTCGAACATTCATGACACATGTGTG 470
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
375 TGAGTGGGATGAGCTGCCCATTTTGGCTTGGCCAGCTCTCAGCCCAACATGGAGCTGATGA 434
QY 471 GGGCCCTCTGACAGCAGCGCAGCGGTGTAACACTCTCTTGGATCCCTCGTGGAAA 530
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
435 CAGTGTCTGAGTCCA---CCGGGAATGTACTTCGTCCAAGTGGGTTCGCCGGGGGGA 491
QY 531 CTACATAGAATCTAATCGTGATGACTGCACGGTGTCTTTGATCTATGCTGTGCACCTTAA 590
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
492 CTACATCGCTCCACACGGCAGATGACACCCACACTGATGTAGCGCGTCAACCTGAA 551
QY 591 GAAGTCAGGCTATGCTTCTTTTGTAGTACCAGTATGTGCAACAACATCTCTCTTTCAGTT 650
```

```
Db 552 GCAATCTGCGACCGTTAACTTCGAATCTACTATCCAGACTCCAGCATCATCTTTCAGTT 611
QY 651 CTTTATTCAAATGATCAGTGCAGGAGATGGACACCCACTGACAAAGTGGGTAAACT 710
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
612 TTTTGGTTCAAGATGACCACTGCGCCCAATGCAGATGATCTC---CAGGTGATGAAGAC 668
QY 711 TACAGACAATGGAGATGGGCTCTCATTTCTGTAATGCTGAATCAGGACACAAACATCT 770
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
669 CACAGAAA---AGGATGGGAATTCACAGTGTGGAGCTAAATCGAGGCAATANTGTCTCT 725
QY 771 CTACTGGAGAACTCAGGCATCTTATGGTTCTTAAGGGGTCAAGCCTCTGCTGGTAA 830
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
726 CTATTGGAGAACACAGAGCTTCTCAGTATGGACCAAGTACCAAGCTGTGCTGTGAG 785
QY 831 AAATATCACAATTTGAAGGGTGGCGGTACACATCAGAAATGTTTCTTGAAGCCAGGCAC 890
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
786 AAACATTTGCCATAACAGGGTGGCTTACACTTCAGAAATGCTTCCCTGCAAACTGGCAC 845
QY 891 ATTACGAAACAACACAGGTTTCACTTCACTGCCAGGTGTGTCAGAAACACCTTATTTCTGA 950
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
846 GTATCGACACAGCAGGCTCTCTTCTGCAAACTTTGCCAGCCAACCTCTTATTCARA 905
QY 951 GAAAGAGCCAAAGAAATGTATAAGGTGTA---AGACGACTCTCAATTTTCAGGATCCAG 1007
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
906 TAAAGGAGAAACTTCTTGGCCACAGTGTGACCTGACAAATCTCAGAGAAAGGATCTTTC 965
QY 1008 TGAGTGTACAGAGCGCCCTCCCTGTACCAACAAGAACTATTTCCAGATCCATCTCATG 1067
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
966 TTTCTGTACGTTGCGCCAGCTTTCACAGACAAAGATTTATTTCTACACACACAGCGCTG 1025
QY 1068 TGATGAAGAAAGAAAGACACAGATATGTACAAAGTGTAGAGCCCAAAATCTGCCGGGA 1127
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1026 CGATGCCACGGAGAGACAACTCATGTACAAATGGCCCAAGCCCAAAATCTGTAGCGA 1085
QY 1128 GGATCTCAGATGTATTTAGATTGCCCTTCTTGGAGAGAAAGATTTGTCGCCCTTG 1187
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1086 GGACCTTGAGGGGGCAGTGAAGCTGCTGCTGTGTGTGAAGACCACTGCCACCCCTG 1145
QY 1188 CAACCTGGATTTTATAAANTGGATCATCTTCTGCCATCCCTGCTCTCTGGAACAT 1247
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1146 CAACCCAGCTTCTTCAAAACCAACAACAGACCTGCCAGCCCTGCCATATATGGTCCCTA 1205
QY 1248 TTCAGATGAAACCAAGAAATGTAGACATGTCCAGCAGGAAGCGCTGCACCTTGGCTT 1307
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1206 CTCCAATGG---CTCAGACTGTACCCTGCTGCGAGGACTGACCTGCTGTGGGAT 1262
QY 1308 TGAATATAAATGGTGAATGTCTTCTCGCAACATGAAACTTCTGCTCAATGTGG 1367
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1263 TGAATACAAATGGTGAACACGCTGCCCAACAACATGGAAGACCGCTTCTCAGTGGGAT 1322
QY 1368 GAATTCAAAGTCGGATGGAATGAATGGTGGGAGTGGCTGGAGATCATATCCAGAGTGG 1427
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1323 CAATTCGAGTACAAGGGCATGACAGGCTGGGAGTGGCTGGATGATCATATTTACACAGC 1382
QY 1428 GGCTGGAGCTTCTGACAATGATTTACCTGATCTTAAACTTTGCATATCCAGGATTTAAACC 1487
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1383 TGTGGAGCTTCAGACAATGACTTTCATGATCTCAGTCTGGTGTGCCAGGATTTAGACC 1442
QY 1488 ACCAATCTATGACTGGAGCCAC---GGGTTCTGAACTAGGAAGAAATCAACATTTCTCT 1544
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1443 TCCGAGTCCGTTGATGGCAGACAGAGAATAAAGAGTGGCCAGAAATCACATTTCTCTT 1502
QY 1545 TGAGACCTCTCTGTCAGCTGACTGTGTTTGTACTTTCATGCTGGTGGATTAATAAGAAAAG 1604
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1503 TGAGACCTCTCTGTCGAACTGTGAGCTCTACTTTTCATGCTGGTGGTGTGAATTTCTAGGAC 1562
QY 1605 TACAATGTGTAGAAATCGTGGGTGGAAACCAAGAAAGAAAGCTTACACCCATATCAT 1664
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1563 CAACACTCTGTGGAGACGTGGAAGGTTCCAAAGGCAACAGTCTCTATACCTACATCAT 1622
QY 1665 CTTCAAGAATGCAACTTTTACATTTTACATGGGCTTCCAGAGAACTTAATCAGGGTCAAGA 1724
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Db 1623 TGAGGAGAACTACACGAGCTTACCTGGGCTTCCAGAGAACACATTTTCATGAGGC 1682
QY 1725 TAATAGCGGTTTCATCAATGACATGCTGAAGATTTATTTATTCACAGCCACCAATGACGT 1784
Db 1683 AAGCAGGAAGTACACCAATGACGTTGCCAAGATCTACTCCATCAATGTCCACCAATGTTAT 1742
QY 1785 TGAATGGGTGGGCTCTCATATGCGGTGCCCTGTCGGTTCCTGAACAGTCGGGTTTCATC 1844
Db 1743 GAATGGGTGGGCTCTACTGCGCTGCCCTGAGAGGCTCTGATGTGGGCTCCTC 1802
QY 1845 GTGTGTCCTCCCTCCAGGCACTACATTTGAGAAGAAACCAACCACTGCAAGGAATG 1904
Db 1803 CTGACCTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1862
QY 1905 TCCACCTGACACCTACTCTGCTATACATCAGGCTTATGCAAGAGGCTGTTGATTCATG 1964
Db 1863 CCCCCCTAACACANTTCTGAAGCCCAACAGCCTTATGGTGTCCAGGCTGTGTGCCCTG 1922
QY 1965 CGGGCTGGGAGTAAACAATCAGGACCACTTCGGTTCCTATGCTGCTGCTGCTGCTGCTGCT 2024
Db 1923 TGGTCCAGGAGCAAGAACCAAGATCCACTCTCTGTCTACAAATGATTCACCTTCTC 1982
QY 2025 CCATGAAAAGAAATCAGATTTGTCATATGACATTTAGCAACCTCAGCAGTGTGGGCTC 2084
Db 1983 AGCAACACTCCCAACAGGACTTTCAACTACAACTTCTCCGCTTTGGCAACACACCGCTAC 2042
QY 2085 ATTAATGAATGGGCCCCAGCTTCACCTCCAAAGGAACAAATACTTCCATTTCTTCAATAT 2144
Db 2043 TCTTGTGGGGCCCAAGCTTCACTTCCAAAGGTTGAATACTTCCATCACTTTACCTT 2102
QY 2145 CAGTTTATGTTGGGATGAGGGAAGAGATGCTCTCTTACCAACATATTAACAGACTT 2204
Db 2103 CAGTCTCTGTGGAACACAGGAGTGAAGAAATGCTGTGTCCAGCAATGTCACAGCT 2162
QY 2205 TACAGTAAAGAAATAGTGGCAGGCTCAGATGATACAAATTTGGTAGGGGCACTTGT 2264
Db 2163 CGGATTCCTGAGGGTGAAGTCAAGGTT-----CTCCAAATCTATCACAGCTTACGT 2213
QY 2265 ATGCCAGTCAACAAATTTATTCCTTCTGAAAGTAAAGGTTTCCGAGCAGCCTTATCATCACA 2324
Db 2214 CTGCCAGGAGTATCATCCGCCAGAGTGACAGCTACAAAGCCGGGTTTCTCTACA 2273
QY 2325 ATCCATCATTCGGCAGATACATTCATAGAGTACAGTTGAAACCAATTTGAAAATAT 2384
Db 2274 GCTGTGACGCTTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2333
QY 2385 TAATAAAGAAAGATATCTTCCAGTTTCCAAAGCCAAATACCAAGATGTCATTTCTT 2444
Db 2334 CACCTCCAGCTGAATTTTCCAGTGGAGTCTTGGGAATACCGGAGCTGATCTTCTT 2393
QY 2445 TTATAAGTCTTCTACAGCAACACATCTTTATTTATGGCCGATCACTGCTGTGAAAAT 2504
Db 2394 TTATAGTCCCAATGATGTACCCAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2453
QY 2505 GAGGTGTAATCTTAATCTGGAGCAGAGTATTTTCAAGTCCCAAGCAAGTGGCCAGC 2564
Db 2454 CAGGTGCACTCCACAGAAACTGTCCCTGGAAGTTTGTCTGCCAGGAACGTGCTCAGA 2513
QY 2565 AGGTACTGTGAGTGGTGTACGTTCTTATTTCTGTTGGGAGTGTGCTGAGCTTGGCCTCT 2624
Db 2514 TGGGACCTGTGATGGTGTCACTTCTGTTGGGAGAGCGGCTGCTTGGCCGCT 2573
QY 2625 GTGTAGGAGCATGACTTCCATGAGATTCAGGGAGCCTGCAAGAGAGGATTTTCAGGAAC 2684
Db 2574 CTGTCAGTGGGTGACTACCGTGTATCTGTCAGCAGCTGTGTGGTGGGATCCAGAAGAC 2633
QY 2685 CTTGTATGTGGAATGAACCTTAATGTTGCTTAAAGGAATTTCTTCCCTCAGAAAAA 2744
Db 2634 TACTTACGTGTGGCGAAGACCAAGCTATGCTCTGCTGGTGCATTTCTCTCCCTGAGCAGAG 2693
QY 2745 GTTGGCAACCTGTGAACCGTTGACTTTTGGCTGAAGGTG 2784
Db 2694 AGTCACCATCTGCAAAACCATAGATTTCTGGCTGAAAGTG 2733

RESULT 11
AAF27997
ID AAF27997 standard: DNA; 2554 BP.
XX
AC AAF27997;
XX
DT 08-MAY-2001 (first entry)
XX
DE Human TR13 receptor coding sequence SEQ ID NO: 1.
XX
KW Human; tumour necrosis factor receptor; TR13; TR14; infection;
KW cancer; autoimmune disease; allergy; inflammatory disease;
KW graft rejection; apoptosis; cardiovascular disease; aneurysm; ds.
XX
OS Homo sapiens.
XX
PN WO200105834-A1.
XX
PD 25-JAN-2001.
XX
PF 14-JUL-2000; 2000WO-US19343.
XX
PR 16-JUL-1999; 99US-0144087.
PR 18-AUG-1999; 99US-0149450.
PR 20-AUG-1999; 99US-0149712.
PR 10-SEP-1999; 99US-0153089.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Ni J, Young PE;
XX
DR WPI: 2001-112682/12.
DR P-PSDB; AAB35328.
XX
PT Nucleic acids encoding 2 human tumor necrosis factor receptor
PT polypeptides (TR13) and (TR14), useful for the prevention, diagnosis
PT and treatment of, e.g. cancers, acquired immune deficiency syndrome and
PT hypohidrotic ectodermal dysplasia.
XX
PS Claim 2: Page 366-369; 418pp; English.
XX
CC The present invention provides the protein and coding sequences of the
CC human tumour necrosis factor receptors TR13 and TR14. These sequences are
CC useful in the diagnosis and treatment of many diseases, including cancer,
CC autoimmune diseases, cardiovascular disorders, allergies,
CC neurodegenerative diseases, graft rejection, inflammation, aneurysms and
CC infections.
XX
SQ Sequence 2554 BP; 642 A; 722 C; 588 G; 602 T; 0 other;
Query Match 21.4%; Score 761.4; DB 22; Length 2554;
Best Local Similarity 60.2%; Pred. No. 1.8e-189;
Matches 1376; Conservative 0; Mismatches 891; Indels 20; Gaps 6;
QY 785 CAGGCATCTTATGGTCTTCAAGCGGTCTGCTGTTAAATAATATCACAAATTG 844
Db 18 CAGCCTCTCAGTATGACCAAGTACCAAGCCTGCTGCTGAGAACATTTGCCATAA 77
QY 845 AAGGGTGGCTGACATCAGAAATGTTTTCCTTGAAGCCAGGCAATTCAGCAACAAC 904
Db 78 CAGGGTGGCTTACACTTTCAGAAATGTTTCCCTGCAAGCTGCGAGTATGAGACAAAGC 137
QY 905 CAGGTTTCATTCACCTGCCAGGTGTGTCAGAAACACCTTATCTGAAAGAGGACCAAG 964
Db 138 AGGCTCTCTTCTGCAAACTTTG-CCAGCAACTCTTATTCAAATAAGGAGAAACTTT 196
QY 965 AATGTATAAGGTGTA--AGACGACTCTCAATTTTCAGGATCCAGTGTGTACAGAGC 1021
Db 197 CTTGCCACAGCTGTGACCCCTGACAAATCTCAGAGAAAGATCTTCTTCTCTGTAACTGC 256
QY 1022 GGCCTCCCTGTACCACAAAGACTATTTCCAGATCCATCTCCATGTGTGTAAGAAAGAA 1081

DT 17-DEC-2001 (first entry)
XX Human cDNA encoding a novel foetal antigen, SEQ ID No 689.
DE
XX Human; foetal tissue antigen; ss; antiinflammatory; neuroprotective;
KW immunomodulator; cardiovascular; cytosolic; nephrotoxic;
KW hyperproliferative disorder; breast neoplasm; cancer;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
XX wound healing; epithelial cell proliferation; food additive.
OS Homo sapiens.
XX
PN WO200155312-A2.
PD
XX 02-AUG-2001.
PF 17-JAN-2001; 2001WO-US01321.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234224.
PR 23-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0234984.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.

PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-488782/53.
DR P-PSDB; AAU21345.
XX
PT New polynucleotides and polypeptides for diagnosing, treating,
PT preventing or prognosing e.g. diseases or disorders of the nervous,
PT musculoskeletal, excretory, gastrointestinal, reproductive, and
PT respiratory systems
XX
XX Claim 1; SEQ ID No 689; 642pp; English.
XX
CC The invention relates to novel nucleic acids encoding novel human foetal
CC antigens. The nucleic acids and proteins are used to prevent, treat (e.g.
CC by gene therapy) or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility
CC to a pathological condition. The antibodies to the antigens can also
CC be used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. Numerous
CC examples of diseases and disorders treated by the nucleic acids and
CC proteins are given in the specification. The present sequence

Query Match 18.1%; Score 645.4; DB 22; Length 672;
Best Local Similarity 98.4%; Pred. No. 2.7e-159;
Matches 662; Conservative 9; Mismatches 0; Indels 2; Gaps 2;

QY 2668 CAAATATTCCAACTAGTACTAGGAGTAACTCAAAAGAGTGTGAAGTCCCGGTGCGAGA 2927
DB 2 CAAATATTCCAACTAGTACTAGGAGTAACTCAAAAGAGTGTGAAGTCCCGGTGCGAGA 61
QY 2928 CAGTTGTGCTATCATGGAAGGAGAGTAATGAAGAGGAAGTTGTATATATTTCCAAATAACA 2987
DB 62 CAGTTGTGCTATCATGGAAGGAGAGTAATGAAGAGGAAGTTGTATATTTCCAAATAACA 121
QY 2988 GTCACTACTAGGAACCTCAATCTTTGGCAACCAAGAAAAGAACCACTTTTGAATC 3047
DB 122 GTCACTACTAGGAACCTCAATCTTTGGCAACCAAGAAAAGAACCACTTTTGAATC 181
QY 3048 TGTTCACACTGAAACCTCAAGATCCCAATATATATGAAGAGACAGTCTGTAGCCTTTGAG 3107
DB 182 TGTTCACACTGAAACCTCAAGATCCCAATATATATGAAGAGACAG-KCTGTAGCCTTTGAG 240
QY 3108 ACTAATGAACAAAGAAACCTGCTCTAGTCTTTTACAGGACCATATTTTAGGGTCTGTCTCTCA 3167
DB 1

Db 241 ACTAATGAACAAAGAAACCTGCTCTAGTCTTTTACAGGACCATATTTTAGGGTCTGTCTCTCA 300
QY 3168 TACCTGTACATTGGTGTATCTACAGAGAGAGGGCCATGCCCTGAAAAGGAGAGATT 3227
Db 301 TACCTG-CACATTGGTGTATCTACAGAGAGAGGGCCATGCCCTGAAAAGGAGAGATT 359
QY 3228 GAAACATTGTGCTTATCATCATGGTCAAGTACCTTGCCTCAAAATAAAGAAAGCAAAATG 3287
Db 360 GAAACATTGTGCTTATCATCATGGTCAAGTACCTTGCCTCAAAATAAAGAAAGCAAAATG 419
QY 3288 ATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAGAGAGAGATTTATCTGTATATACACA 3347
Db 420 ATTTGGGTCTCAACTGAAGATGAAGCTCAACTCAGAGAGAGATTTATCTGTATATACACA 479
QY 3348 TAACTGAACAAAGTTTAAAGCCCAACCATGACATGCTGCTCATGATGCCCATATAAATAATG 3407
Db 480 TAACTGAACAAAGTTTAAAGCCCAACCATGACATGCTGCTCATGATGCCCATATAAATAATG 539
QY 3408 GGTAACTTTTATCTTATGATGTCTACATAACAAGTGTGATTTGGAAGGCACATGTGAG 3467
Db 540 GGTAACTTTTATCTTATGATGTCTACATAACAAGTGTGATTTGGAAGGCACATGTGAG 599
QY 3468 CATATGCATTATCATGCCAAATTTATGTTTTCTTTCTTTTATATATTTTGGGGGAAAAATAAA 3527
Db 600 CATATGCATTATCATGCCAAATTTATGTTTTCTTTCTTTTATATATTTTGGGGGAAAAATAAA 659
QY 3528 TTTTAAAGGTA 3540
Db 660 TTTTAAAGGTA 672
RESULT 13
ABL69028/c
ID ABL69028 standard: DNA; 576 BP.
XX
AC ABL69028;
XX
DT 15-MAY-2002 (first entry)
XX
DE Kidney cancer-related gene sequence SEQ ID NO:7365.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US10838.
XX
PR 05-JUN-2000; 2000US-209473P.
PR 05-JUN-2000; 2000US-209531P.
PR 18-SEP-2000; 2000US-231133P.
PR 18-SEP-2000; 2000US-233617P.
PR 20-SEP-2000; 2000US-234009P.
PR 20-SEP-2000; 2000US-234034P.
PR 22-SEP-2000; 2000US-234052P.
PR 22-SEP-2000; 2000US-234509P.
PR 22-SEP-2000; 2000US-234567P.
PR 25-SEP-2000; 2000US-234923P.
PR 25-SEP-2000; 2000US-234944P.
PR 25-SEP-2000; 2000US-235077P.
PR 25-SEP-2000; 2000US-235082P.
PR 25-SEP-2000; 2000US-235134P.
PR 25-SEP-2000; 2000US-235280P.
PR 26-SEP-2000; 2000US-235637P.
PR 26-SEP-2000; 2000US-235638P.
PR 27-SEP-2000; 2000US-235711P.
PR 27-SEP-2000; 2000US-235720P.

```
PR 27-SEP-2000; 2000US-235840P.
PR 27-SEP-2000; 2000US-235863P.
PR 28-SEP-2000; 2000US-236028P.
PR 28-SEP-2000; 2000US-236032P.
PR 28-SEP-2000; 2000US-236033P.
PR 28-SEP-2000; 2000US-236034P.
PR 28-SEP-2000; 2000US-236109P.
PR 28-SEP-2000; 2000US-236111P.
PR 29-SEP-2000; 2000US-236842P.
PR 29-SEP-2000; 2000US-236891P.
PR 02-OCT-2000; 2000US-237172P.
PR 02-OCT-2000; 2000US-237173P.
PR 02-OCT-2000; 2000US-237278P.
PR 02-OCT-2000; 2000US-237294P.
PR 02-OCT-2000; 2000US-237295P.
PR 02-OCT-2000; 2000US-237316P.
PR 03-OCT-2000; 2000US-237425P.
PR 03-OCT-2000; 2000US-237598P.
PR 03-OCT-2000; 2000US-237604P.
PR 03-OCT-2000; 2000US-237606P.
PR 03-OCT-2000; 2000US-237608P.
PR 01-NOV-2000; 2000US-244867P.
PR 01-NOV-2000; 2000US-245084P.
XX
PA (AVAL-) AVALON PHARM.
XX
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX
XX WPI; 2002-188264/24.
XX
XX Screening for anti-neoplastic agent involves exposing cells to a
PT chemical agent to be tested for anti-neoplastic activity, and
PT determining a change in expression of a gene of a signature gene set
XX
XX Claim 1; SEQ ID 7365; 44pp; English.
XX
XX The present invention describes a method (M1) for screening for an
CC anti-neoplastic agent. The method involves exposing cells to a chemical
CC agent to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABU61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening
CC an anti-neoplastic agent, and can be used for producing a product which
CC is the data collected with respect to the anti-neoplastic agent as a
CC result of M1, and the data is sufficient to convey the chemical
CC structure and/or properties of the agent. M1 can be used in the
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC carcinoma, papillary carcinoma and Wilms' tumour.
XX
XX Query Match 14.2%; Score 504; DB 24; Length 576;
Best Local Similarity 94.4%; Pred. No. 4.1e-122;
Matches 543; Conservative 0; Mismatches 28; Indels 4; Gaps 2;
QY 2970 TGTATATCCATAAAGCTCACTACTAGGAAACCTCAATCTTTGGCAACCAAGGAAAA 3029
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
575 TGTGTTTTCCATTAAACCGCTCACTACTNGAACCCTCAANTCTTTGGGCAACCCAGGGAA 516
QY 3030 AGNAGA---CCATTTTGATCTGTT--CAACTGAAACCTCAAGATCCCCAATATATGAA 3085
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
515 AAAGGAGACCCATTTTGATCTGTTCCAACTGAAACCCCTCAAGATCCCCAATATATGAA 456
QY 3086 GAGACAGTCTGTAGCTTTGAGACTAATGCAACAAGAACCTGCTCTAGTTTACAGAC 3145
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
455 GAGACAGTCTGTAGCTTTGAGACTAATGCAACAAGAACCTGCTCTAGTTTACAGGAC 396
```

```
QY 3146 CATATTTTAGGGTCTGTCCTCATACCTGTGCACATTTGGTGATCTCACAGAGAGGGCCATG 3205
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
395 CATATTTTAGGGTCTGTCCTCATACCTGTGCACATTTGGTGATCTCACAGAGAGGGCCATG 336
QY 3206 CCGCTGAAAAAGGAGAGATTGAAACATTTGATTTGCCCTTTATCACATTTGGTCAAGTACCTT 3265
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
335 CCGCTGAAAAAGGAGAGATTGAAACATTTGATTTGCCCTTTATCACATTTGGTCAAGTACCTT 276
QY 3266 GCCAATTAAGGAAAGCAATGATTTGGGTCCTCACTGAAGATGAAGCTCAACTCAGGAA 3325
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
275 GCCAATTAAGGAAAGCAATGATTTGGGTCCTCACTGAAGATGAAGCTCAACTCAGGAA 216
QY 3326 GAGATTTATCTGTATATACACATAACTGAAACCAAGTTTAAAGCTCAACTGCT 3385
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
215 GAGATTTATCTGTATATACACATAACTGAAACCAAGTTTAAAGCTCAACTGCT 156
QY 3386 GATGCATGCCATATATTAATGGGTAACTTTTATTTCTTTATGATGCTCTACATAACAAGTG 3445
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
155 GATGCATGCCATATATTAATGGGTAACTTTTATTTCTTTATGATGCTCTACATAACAAGTG 96
QY 3446 TGATTTGGAGGCACATGTCAGCATATGATGATCAATTTATGTTTCTTTCTTTCTTTGTT 3505
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
95 TGATTTGGAGGCACATGTCAGCATATGATGATCAATTTATGTTTCTTTCTTTCTTTGTT 36
QY 3506 TATATTTTGGGAAAAATTAATAATTTTAAAGTA 3540
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
35 TATATTTTGGGAAAAATTAATAATTTTAAAGTA 1
RESULT 14
AAAX27340
ID AAX27340 standard; DNA; 1129 BP.
XX
AC AAX27340;
XX
DT 11-JUN-1999 (first entry)
XX
XX Human secreted protein gene 30 clone H0DAZ26.
XX
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
immune system; ischaemic shock; lymphocytic disease; brain; hepatic; lymphoma;
inflammation; cognitive disorder; schizophrenia; Alzheimer's disease; testis; AIDS;
osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS Homo sapiens.
XX
XX WO9902546-A1.
XX
XX 21-JAN-1999.
XX
XX 07-JUL-1998; 98WO-US13684.
XX
PR 12-SEP-1997; 97US-0058785.
PR 08-JUL-1997; 97US-0051916.
PR 08-JUL-1997; 97US-0051918.
PR 08-JUL-1997; 97US-0051919.
PR 08-JUL-1997; 97US-0051920.
PR 08-JUL-1997; 97US-0051925.
PR 08-JUL-1997; 97US-0051926.
PR 08-JUL-1997; 97US-0051928.
PR 08-JUL-1997; 97US-0051929.
PR 08-JUL-1997; 97US-0051930.
PR 08-JUL-1997; 97US-0051931.
PR 08-JUL-1997; 97US-0051932.
PR 08-JUL-1997; 97US-0052732.
PR 08-JUL-1997; 97US-0052733.
PR 08-JUL-1997; 97US-0052793.
PR 08-JUL-1997; 97US-0052795.
PR 08-JUL-1997; 97US-0052803.
```


Matches 857: Conservative 0: Mismatches 585: Indels 9: Gaps 1:

QY 1615 GTAGATCGTGGGCTGGACCAAGAAACAAAGCTTACACCCATATCATCTTCAAGAAT 1674
Db 21 GTGGAGACGTGGAAGGTTTCCAAAGGCAACAGCTTATACCTACATCATTTGAGGAGAAC 80
QY 1675 GCAACTTTTACATTTACATGGGCAATTCAGAGAACTAATCAGGGTCAAGATAATAGACGG 1734
Db 81 ACTACGACGAGCTTCACCTGGGCTTCAGAGAGACCACTTTTCATGAGGCAAGCAGGAAG 140
QY 1735 TTCTCAATGACATGGTGAAGATTTATTTCTATACAGCCCACTAATGCAAGTGTGAGGGTG 1794
Db 141 TACACCAATGAGTGGCAAGATCTACTCCATCAATGTCCACCAATGTTATGAATGGGGTG 200
QY 1795 GCGTCTCATGCCGTGGCTGTGCCCTCGGTTCTGAACAGTCGGGTTTCATCGTGTGTCGCC 1854
Db 201 GCTCTCTACTGCCCTGCCCTGTGCCCTAGAGCCCTCTCATGTGGGCTCCTCCTGCACCTCT 260
QY 1855 TGCCCTCCAGGCCACTACATTTGAGAAAGAACCAACAGCTGCAAGGAATGTCCACCTGAC 1914
Db 261 TGTCTGTGTTACTATATTGACCGAGATTTCAGGAACCTGCCACTCCTGCCCCCTTAAC 320
QY 1915 ACCTACCTGTCCATFACATCAGGCTTATGGCAAGAGGCTTGTATTCCATCGGGCCTGGG 1974
Db 321 ACAATCTGAAAGCCCAAGCCTTATGGTGTCCAGGCCGTGTGCCCTGTGGTCCAGGG 380
QY 1975 AGTAAACAAATCAGGACCAATTCGGTTTGGTATAGTGAAGTCTTTTCTACCATGAAAA 2034
Db 381 ACCAAGAACAAACAAAGATCCACTCTCTGTGTGACAAATGTCACCTTCTCACGCAACACT 440
QY 2035 GAAATCAGATTTTGGCACTATGACTTTAGCAACCTCAGCAGTGTGGCTCATTAATGAAT 2094
Db 441 CCAACGAGGACTTCAACTACACTCTCCGCTTGGCAACACCGCTCACTCTTGTGGA 500
QY 2095 GGCCCCAGCTTCACTCCAAAGGAACAAATACTTTCATTTCTTCAATATCAGTTTATGT 2154
Db 501 GGGCAAGCTTCACTTCCAAAGGTTGAATACTTCCATCCTTACCCTCAGTCTCTGT 560
QY 2155 GGGCATGAGGGAGAGATGGCTCTCTGTACCACAAATATAACAGACTTTACAGTAAAA 2214
Db 561 GGAACCAAGGTTAGGAAATGTGTGTGACCGGACAAATGTCACTGACCTCCCGGATTCCT 620
QY 2215 GAAATAGTGGCAGGGTCAGATGATTACACAAATTTGGTAGGGCACTTGTATGCCAGTCA 2274
Db 621 GAGGGTGAAGTCAGG-----TTCTCCAATCTATCACGCCCTACGCTGCCAGGCA 671
QY 2275 ACAATATTCCCTTCTGAAAGTAAAGGTTTCCGAGCAGCCTTATCATCACAAATCCATCT 2334
Db 672 GTCATCATCCCCCAGAGGTGACAGGCTACAGGCGGGGTTTCCCTCACAGCCTGTGACG 731
QY 2335 CTGGCAGATACATTATAGGAGTCACAGTTGAAACCACTTGAANAATATTAAATATAAAA 2394
Db 732 CTTCGTGATCGACTTATTTGGGTGACACAGATATGACTCTGGATGGAATCACCTCCCCA 791
QY 2395 GAAGATATGTTCCAGTTCCTCAACAGCCAAATACCAAGATGTGCAATTTCTTTTATAGTCT 2454
Db 792 GCTGAACCTTTTCCACCTGGAGTCTTGGGAATACCGGACGTGATCTCTTTTATAGTCC 851
QY 2455 TCTACAGCAACACACTTCTGTATTATATGGCCGATCAACTGCTGTGAAAATAGGTTGAAT 2514
Db 852 ATGATGTGACCCAGCTCTCTGTCAGTTCTGGGAGATCAACCCACCATCGCGCTGAGTGCAGT 911
QY 2515 CTTACTAAATCTGGAGCAGGAGTGTATTTTCAGTCCCGCAGCAAGTGCACGAGTACCTGT 2574
Db 912 CCACAGAAAACCTGTCCCTGGAAGTTTGTCTGCGCAGGAACGTGTGATGAGGACCTGT 971
QY 2575 GATGGGTGTACGTTCTATTCTCTGGGAGAGTGTGAAGCTTGGCCCTCTGTGTACGGAG 2634
Db 972 GATGGGTGCAACTTCCACTTCTCTGTGGGAGCGGGTGTGTGCCGCTCTGCTCAGTG 1031
QY 2635 CATGACTTCCATGAGATTGAGGAGGCTCAGAGAGGATTTTCAGGAAACCTTGTATGTG 2694
Db 1032 GCTGACTACCATGCTATCTCAGCAGCTGTGTGGGTGGGATCCAGAAAGACTTACTTACGTG 1091

Search completed: May 12, 2003, 00:39:48
Job time : 738.514 secs

BACKFILE DOCUMENT INDEX SHEET

REV 7/02/03

APPL PARTS

A DOCPHOENIX

IMIS	Internal Misc. Paper
LET.	Misc. Incoming Letter

PCT Papers in a 371 Application
371P
A...
Amendment Including Elections
ABST
ADS
Application Data Sheet
AF/D
Affidavit or Exhibit Received
APPENDIX
ARTIFACT
Artifact
BIB
Bib Data Sheet
CLM
Claim
COMPUTER
Computer Program Listing
CRFL
All CRF Papers for Backfile
DIST
Terminal Disclaimer Filed
DRW
Drawings
FOR
Foreign Reference
FRPR
Foreign Priority Papers
IDS
IDS Including 1449

Internal	Examiner Search Notes	SRNT	CLMPTO	PTO Prepared Complete Claim Set	
ECBOX	Evidence Copy Box Identification	WCLM	Claim Worksheet	WFEF	Fee Worksheet

CTMS	Misc. Office Action
1449	Signed 1449
892	Abandonment
ABN	Board of Appeals Decision
APDEC	Examiner Answer
CTAV	Count Advisory Action
CTEQ	Count Ex parte Quayle
CTFR	Count Final Rejection

OUTGOING

NPL
Non-Patent Literature
OATH
Oath or Declaration
PET.
Petition
RETMAL
Mail Returned by USPS
SEQLIST
Sequence Listing
SPEC
Specification
SPEC NO
Specification Not in English
TRNA
Transmittal New Application

CTNF
Count Non-Final
CTRS
Count Restriction
EXIN
Examiner Interview
M903
DO/EO Acceptance
M905
DO/EO Missing Requirement
NFDR
Formal Drawing Required
NOA
Notice of Allowance
PETDEC
Petition Decision

INCOMING

AP.B
Appeal Brief
C.AD
Change of Address
N/AP
Notice of Appeal
PA..
Change in Power of Attorney
REM
Applicant Remarks in Amendment
XT/
Extension of Time filed separate

File Wrapper

FWCLM	File Wrapper Claim
IIFW	File Wrapper Issue Information
SRFW	File Wrapper Search Info



